

# U.F.R. Economie Appliquée

Maîtrise d'Economie Appliquée

Cours de Tronc Commun

## Econométrie Appliquée Séries Temporelles

Christophe HURLIN

## Chapitre 2

### Tests de Non Stationnarité et Processus Aléatoires Non Stationnaires

Dans le premier chapitre, nous avons vu qu'une des première étape de la démarche de modélisation d'une série temporelle consiste à vérifier la stationnarité du processus générateur de données. Généralement, on se limite à vérifier la *stationnarité faible* ou *stationnarité du second ordre*.

Nous allons à présent étudier de façon de plus précise ce qu'est un processus non stationnaire. Il existe en effet deux sorte de non stationnarité : la non stationnarité déterministe et la non stationnarité stochastique. Nous verrons que suivant l'origine de la non stationnarité, il convient d'adopter une méthode de stationnarisation particulière.

La seconde partie de ce chapitre sera ensuite consacrée à la présentation des principaux tests de non stationnarité. Il s'agit alors de définir une stratégie empirique permettant de vérifier si les processus sont stationnaires ou au contraire si il est nécessaire de les stationnariser et quelle est alors la méthode appropriée.

## 1 Processus non stationnaires

Dans le premier chapitre, nous avons introduit la notion de stationnarité du *second ordre* ou *stationnarité faible*. D'après cette définition, *un processus est stationnaire au second ordre si l'ensemble de ses moments d'ordre un et d'ordre deux sont indépendants du temps*. Par opposition, *un processus non stationnaire est un processus qui ne satisfait pas l'une ou l'autre de ces deux conditions*. Ainsi, l'origine de la non stationnarité peut provenir d'une dépendance du moment d'ordre un (l'espérance) par rapport au temps et/ou d'une dépendance de la variance ou des autocovariances par rapport au temps.

Le fait qu'un processus soit stationnaire ou non conditionne le choix de la modélisation que l'on doit adopter. En règle générale, si l'on s'en tient notamment à la méthodologie de Box et Jenkins, si la série étudiée est issue d'un processus stationnaire, on cherche alors le meilleur modèle parmi la classe des processus stationnaire pour la représenter, puis on estime ce modèle. En revanche *si la série est issue d'un processus non stationnaire, on doit avant toutes choses, chercher à la "stationnariser", c'est à dire trouver une transformation stationnaire de ce processus*. Puis, on modélise et l'on estime les paramètres associés à la composante stationnaire.

*La difficulté réside dans le fait qu'il existe différentes sources de non stationnarité et qu'à chaque origine de la non stationnarité est associée une méthode propre de stationnarisation*. Nous allons donc commencer dans cette section par présenter deux classes de processus non stationnaires, selon la terminologie de Nelson et Plosser (1982) : les processus *TS* (*Time Stationary*) et les processus *DS* (*Differency Stationary*). Dans la section suivante, nous présenterons les méthodes de stationnarisation pour chacune de ces classes de processus. Mais au delà des enjeux de modélisation économétriques, nous verrons dans cette partie, que l'origine de la non stationnarité a de très fortes

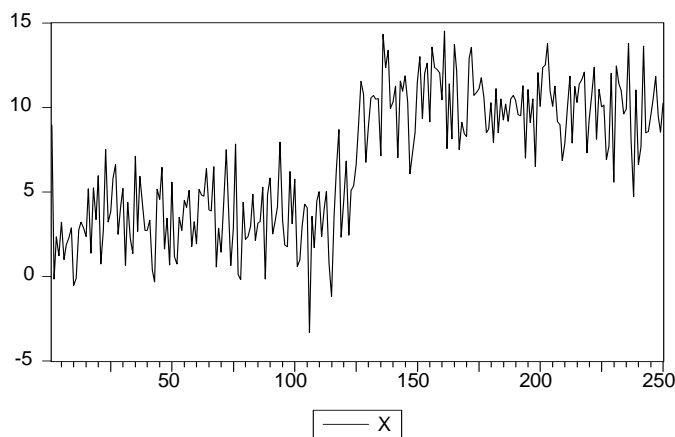
implications sur l'analyse économique des séries que l'on étudie. Nous verrons en particulier que *pour les processus DS il existe une propriété de persistance des chocs qui n'existe pas dans les processus TS*. Une telle hypothèse implique par exemple que si les séries macroéconomiques satisfont une représentation de type *DS*, l'impact des chocs conjoncturels peut avoir un effet permanent sur le niveau de la série étudiée.

Avant de présenter de façon formelle les différentes sources de non stationnarité, nous allons considérer quelques exemples simples de processus non stationnaires. Rappelons au passage la définition de la stationnarité du second ordre (*cf.* chapitre 1) :

**Definition 1** *Un processus  $(x_t, t \in \mathbb{Z})$  est dit stationnaire au second ordre, ou stationnaire au sens faible, ou stationnaire d'ordre deux si les trois conditions suivantes sont satisfaites :*

- $\forall t \in \mathbb{Z}, E(x_t^2) < \infty$
- $\forall t \in \mathbb{Z}, E(x_t) = m$ , **indépendant de  $t$**
- $\forall (t, h) \in \mathbb{Z}^2, cov(x_t, x_{t+h}) = E[(x_{t+h} - m)(x_t - m)] = \gamma(h)$ , **indépendant de  $t$**

Figure 1.1: Processus Non Stationnaire : Modèle avec Rupture



Sur la figure (1.1), est représentée une simulation d'un processus  $(x_t, t \in \mathbb{Z})$  présentant une rupture de moyenne à partir de la date  $t_0 = 125$ , avec  $(\varepsilon_t, t \in \mathbb{Z})$  *i.i.d.*  $(0, 2)$  :

$$\begin{aligned} x_t &= 3 + \varepsilon_t & t < t_0 \\ x_t &= 10 + \varepsilon_t & t \geq t_0 \end{aligned}$$

Ce processus est par définition non stationnaire (condition sur le moment d'ordre un) et l'on vérifie bien sur la figure (1.1) que la réalisation de la moyenne empirique (estimateur convergent de

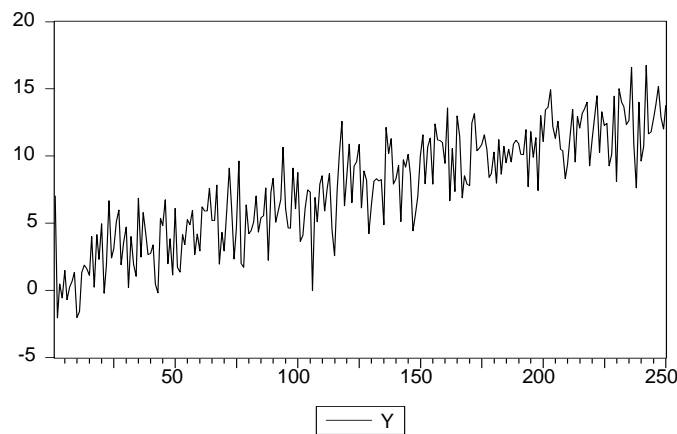
l'espérance) dépendra alors de l'échantillon considéré (avant  $t_0$ , ou après  $t_0$ ). Dans ce cas, la non stationnarité provient de la rupture du modèle, ou plus généralement de la non linéarité.

Un autre exemple de processus non stationnaire est celui du processus suivant :

$$x_t = 1 + 0.05t + \varepsilon_t$$

avec  $\varepsilon_t i.i.d.N(0, 1)$ . Dans ce cas, le processus  $x_t$  correspond à la somme d'une fonction linéaire du temps,  $f(t) = 1 + 0.05t$  et d'un bruit blanc.

Figure 1.2: Processus Non Stationnaire : Trend Déterministe



On voit clairement sur le graphique (1.2) que ce processus ne satisfait pas la seconde condition de la définition de la stationnarité du second ordre. En effet  $E(x_t) = 1 + 0.05t$  croît avec le temps, à chaque date la variable aléatoire  $x_t, \forall t \in \mathbb{Z}$  a une espérance plus grande que celle de  $x_{t-1}, x_{t-2}, \dots, x_{t-j} \dots$ . Dans ce cas, l'origine de la non stationnarité provient tout naturellement de l'inclusion de la tendance (ou plus généralement de la fonction du temps  $f(\cdot)$ ) dans la définition du processus  $(x_t, t \in \mathbb{Z})$ . On dit que la non stationnarité est alors de *type déterministe*.

Mais il existe d'autres sources de non stationnarité. Considérons le processus suivant, que l'on qualifie généralement de *marche aléatoire pure (Random Walk Process)* ou *marche aléatoire sans dérive* :

$$x_t = x_{t-1} + \varepsilon_t \tag{1.1}$$

avec  $\varepsilon_t i.i.d.N(0, \sigma_\varepsilon^2)$ . A priori dans ce cas, la non stationnarité n'est pas de type déterministe, puisque le processus  $x_t$  ne comporte pas de fonction déterministe du temps (figure 1.3). Pourtant ce type de processus est aussi non stationnaire. Cherchons donc à déterminer l'origine de cette non

stationnarité. Le processus  $(x_t, t \in \mathbb{Z})$  peut se réécrire sous la forme :

$$\begin{aligned} x_t &= \varepsilon_t + \varepsilon_{t-1} + \varepsilon_{t-2} + \dots + \varepsilon_{t-j} + \dots \\ &= \sum_{j=0}^{\infty} \varepsilon_{t-j} \end{aligned} \quad (1.2)$$

Dès lors, connaissant les propriétés du bruit blanc  $\varepsilon_t$ , on montre que :

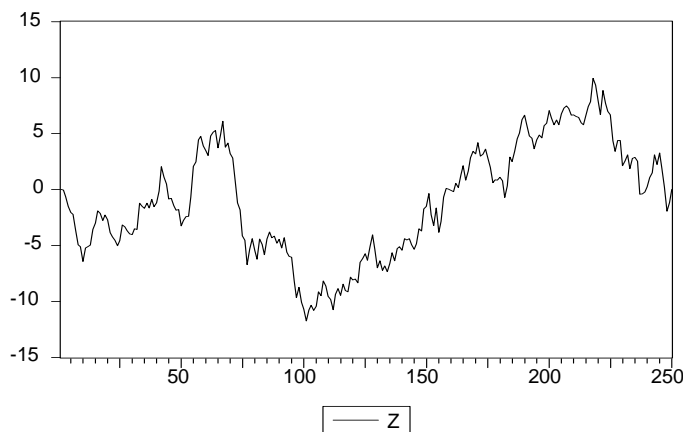
$$E(x_t) = E\left(\sum_{j=0}^{\infty} \varepsilon_{t-j}\right) = \sum_{j=0}^{\infty} E(\varepsilon_{t-j}) = 0 \quad (1.3)$$

Donc le processus  $x_t, \forall t \in \mathbb{Z}$  a une espérance nulle et donc satisfait la seconde condition de la définition de la stationnarité. Mais il ne satisfait pas la première condition puisque :

$$V(x_t) = V\left(\sum_{j=0}^{\infty} \varepsilon_{t-j}\right) = \sum_{j=0}^{\infty} V(\varepsilon_{t-j}) = \sum_{j=0}^{\infty} \sigma_{\varepsilon}^2 \equiv \infty \quad (1.4)$$

La variance de  $x_t$  est non convergente. De plus, si l'on avait définie une condition initiale  $x_0$ , alors la variance de  $x_t$ , définie par  $V(x_t) = \sum_{j=0}^{t-1} \sigma_{\varepsilon}^2 = t\sigma_{\varepsilon}^2$  serait fonction de  $t$ . La troisième condition de la stationnarité faible est alors violée. Le processus  $(x_t, t \in \mathbb{Z})$  est donc un processus non stationnaire. Pourtant, l'examen d'une réalisation quelconque de ce processus (figure 1.3) ne permet pas a priori de dire que cette variable est non stationnaire. On sent d'ores et déjà, la nécessité de proposer des tests de l'hypothèse de stationnarité.

Figure 1.3: Processus Non Stationnaire : Marche Aléatoire Sans Dérive



Dans ce dernier cas, la non stationnarité du processus  $(x_t, t \in \mathbb{Z})$  tient au fait que les chocs  $\varepsilon_t$  s'accumulent au cours du temps, ce qui accroît la variance de  $x_t$  au fur et à mesure que le temps passe. *L'origine de la non stationnarité provient ici de l'accumulation de chocs stochastiques  $\varepsilon_t$  :*

la non stationnarité peut donc être de type stochastique. Le fait que la stationnarité puisse être de type déterministe ou stochastique nous amène à présent à définir la classe des processus *TS* (*Trend Stationary*), qui correspondent à une non stationnarité de type déterministe et la classe des processus *DS* (*Difference Stationary*), qui correspondent à une non stationnarité de type stochastique. Cette distinction selon l'origine de la non stationnarité est essentielle tant sur le plan statistique que sur le plan de l'analyse économique.

## 1.1 Les processus TS

Commençons par définir ce qu'est un processus *TS* pour *Trend Stationary*, selon la terminologie proposée par Nelson et Plosser (1982)

**Definition 2**  $(x_t, t \in \mathbb{Z})$  est un processus *TS* s'il peut s'écrire sous la forme

$$x_t = f(t) + z_t \quad (1.5)$$

où  $f(t)$  est une fonction du temps et  $z_t$  est un processus stochastique stationnaire.

Dans ce cas, le processus  $x_t$  s'écrit comme la somme d'une fonction déterministe du temps et d'une composante stochastique stationnaire, éventuellement de type *ARMA*. Dès lors, il est évident que le processus ne satisfait plus la définition de la stationnarité du second ordre. En effet, on montre immédiatement que  $E(x_t) = f(t) + z$  où  $z = E(z_t)$ , dépend du temps, ce qui viole la seconde condition de la définition d'un processus stationnaire.

L'exemple le plus simple d'un processus *TS* est celui d'une tendance linéaire perturbée par un bruit blanc. On pose  $f(t) = a_0 + a_1 t$  et  $z_t = \varepsilon_t$  :

$$x_t = a_0 + a_1 t + \varepsilon_t \quad (1.6)$$

avec  $(a_0, a_1) \in \mathbb{R}^2$ ,  $\varepsilon_t$  *i.i.d.*  $(0, \sigma_\varepsilon^2)$ . Dans ce cas, on vérifie que le processus  $x_t$  est non stationnaire puisque l'espérance,  $E(x_t) = a_0 + a_1 t$ , dépend de  $t$ . En revanche, le processus  $y_t$  défini par l'écart entre  $x_t$  et la composante déterministe  $f(t) = a_0 + a_1 t$ , est quand à lui stationnaire :  $y_t = x_t - a_0 + a_1 t = \varepsilon_t$  est un bruit blanc, par définition stationnaire.

Une des propriétés importantes de ce type de processus réside dans l'influence des innovations stochastiques  $\varepsilon_t$ . En effet, nous allons montrer que lorsque un processus *TS* est affecté par un choc stochastique, l'effet de ce choc tend à disparaître au fur et à mesure que le temps passe : c'est la propriété de non persistance des chocs. De façon plus formelle, cette propriété est la suivante :

**Propriété 1. L'influence d'un choc  $\varepsilon_t$  à une date  $T$  sur un processus  $x_t$  défini par**

$$x_t = f(t) + z_t$$

**avec  $z_t$  stationnaire et  $E(z_t) = 0$ , est transitoire. La tendance du modèle étant déterministe, après le choc  $\varepsilon_T$ , la séquence des  $x_t$  converge ainsi vers sa valeur de long terme définie par  $f(t)$ . Il n'y a pas de persistance des chocs.**

Cela signifie que lorsque l'on a un processus  $TS$ , en cas de choc positif ou négatif à une date donnée, toutes choses égales par ailleurs, l'influence de ce choc a tendance à s'estomper au cours du temps. La variable considérée rejoint alors sa dynamique de long terme déterminée par  $f(t)$ . Dans le cas où  $f(t)$  est une fonction affine du temps, la variable rejoint la tendance linéaire de long terme. Cette propriété traduit l'existence d'une tendance non stochastique, et qui donc ne présente pas de rupture dès lors que la fonction  $f(t)$  est continue. Economiquement, cela signifie que la trajectoire de long terme de la série est insensible aux aléas conjoncturels.

Afin d'illustrer cette propriété considérons l'exemple suivant où l'on a introduit une structure autorégressive dans la perturbation  $z_t$ :

$$x_t = a_0 + a_1 t + z_t \quad (1.7)$$

$$z_t = \theta z_{t-1} + \varepsilon_t \quad (1.8)$$

où  $(a_0, a_1) \in \mathbb{R}^2$ ,  $|\theta| < 1$  et  $\varepsilon_t$  *i.i.d.*  $(0, \sigma_\varepsilon^2)$ . Le processus  $z_t$  est un  $AR(1)$  stationnaire, puisque la racine associée à son polynôme autorégressif, égale à  $1/\theta$ , est supérieure à l'unité en module. Admettons que  $E(z_t) = 0$ . Etudions à présent l'influence du choc  $\varepsilon_t$  à une date  $T$  quelconque sur la séquence des  $(x_t, t \geq T)$ . Pour ce faire, appliquons la décomposition de Wold au processus stationnaire  $z_t$ , il vient :

$$z_t = \sum_{j=0}^{\infty} \theta^j \varepsilon_{t-j}$$

On peut alors réécrire le processus  $x_t$  sous la forme suivante :

$$x_t = a_0 + a_1 t + \sum_{j=0}^{\infty} \theta^j \varepsilon_{t-j}$$

Supposons qu'à la date  $T$ , on a ait une réalisation du choc  $\varepsilon_T$  positive ( $\varepsilon_T > 0$ ) et qu'ensuite les chocs  $\varepsilon_t$  pour  $t \geq T$  soient nuls. A la date  $T$ , on a donc :

$$x_T = a_0 + a_1 T + \sum_{j=0}^{\infty} \theta^j \varepsilon_{T-j}$$

A la date  $T + 1$ , les chocs  $\varepsilon_{T+1}$  étant nul, on obtient :

$$x_{T+1} = a_0 + a_1 (T + 1) + \sum_{j=1}^{\infty} \theta^j \varepsilon_{T+1-j}$$



De façon générale, à toute date  $T + k$ ,  $k \geq 0$ , le processus  $x_{T+k}$  est défini par :

$$x_{T+k} = a_0 + a_1(T+k) + \sum_{j=k}^{\infty} \theta^j \varepsilon_{T+k-j} \quad (1.9)$$

On peut alors montrer que plus le temps "passe", c'est à dire plus  $k$  est élevé, plus l'influence du choc  $\varepsilon_T$  de la date  $T$  s'estompe. *L'influence de ce choc est donc transitoire.* En effet, la v.a.r.  $x_{T+k}$  va converger vers sa valeur de long terme définie par la tendance linéaire  $f(\cdot)$ . Pour le montrer, il suffit de considérer l'écart entre  $x_{T+k}$  et la valeur correspondante de la tendance  $f(T+k) = a_0 + a_1(T+k)$  et de montrer que cet écart converge vers 0 quand  $k$  tend vers l'infini. On pose  $\tilde{x}_t = x_t - f(t)$ . A la date  $T+k$ , on a  $\tilde{x}_{T+k} = x_{T+k} - a_0 + a_1(T+k)$  et d'après l'équation (1.9), cet écart à la tendance linéaire, peut se réécrire sous la forme :

$$\tilde{x}_{T+k} = \sum_{j=k}^{\infty} \theta^j \varepsilon_{T+k-j} = \lim_{n \rightarrow \infty} \left( \theta^k \varepsilon_T + \theta^{k+1} \varepsilon_{T-1} + \dots + \theta^{k+n} \varepsilon_{T-n} \right) \quad (1.10)$$

Maintenant, voyons ce qui se passe lorsque l'on s'éloigne de la date du dernier choc  $T$ , c'est à dire lorsque  $k$  tend vers l'infini.

$$\lim_{k \rightarrow \infty} \tilde{x}_{T+k} = \lim_{k \rightarrow \infty} \left[ \lim_{n \rightarrow \infty} \left( \theta^k \varepsilon_T + \theta^{k+1} \varepsilon_{T-1} + \dots + \theta^{k+n} \varepsilon_{T-n} \right) \right] \quad (1.11)$$

Sous l'hypothèse que le processus  $z_t$  est stationnaire, c'est à dire sous l'hypothèse que  $|\theta| < 1$ , alors montre immédiatement que :

$$\tilde{x}_{T+k} \xrightarrow[k \rightarrow \infty]{p} 0 \quad (1.12)$$

Ce résultat signifie que lorsque  $k$  tend vers l'infini, l'écart entre le processus  $x_{T+k}$  et la tendance linéaire, converge en probabilité vers 0. *Ainsi, sous l'effet des chocs antérieurs à la date  $T$ , on s'est éloigné de la tendance, mais lorsque le temps passe, l'effet de ces chocs s'estompe et l'on rejoint la tendance de long terme.* **Ce résultat illustre l'absence de persistance des chocs, ou l'absence d'hystérésis pour les processus  $TS$ .**

Nous allons à présent réaliser une expérience numérique pour illustrer cette propriété de non persistance. On considère le processus  $x_t = 1 + 0.05t + z_t$  avec  $z_t = \theta z_{t-1} + \varepsilon_t$  où  $z_0 = 1$ ,  $\theta = 0.4 < 1$  et  $\varepsilon_t$  *i.i.d.*  $(0, 1)$ . Le profil général d'une réalisation de ce processus est celui de la figure (1.2). Nous allons à présent écrire un petit programme sous Eviews dans lequel nous allons annuler toutes les réalisations du choc  $\varepsilon_t$  à partir de la date  $T = 50$ , puis nous allons comparer les évolutions du processus  $x_t$  et de la tendance déterministe  $f(t) = 1 + 0.05t$ . Le programme<sup>1</sup> utilisé sous Eviews est le suivant :

---

<sup>1</sup>Pour lancer ce programme, il convient au préalable de créer un *Workfile* de type *Undated or Irregular*, de 1 à 100, puis de charger le programme (*open program*) ou de le taper (*new program*), et ensuite de l'exécuter (*Run*).

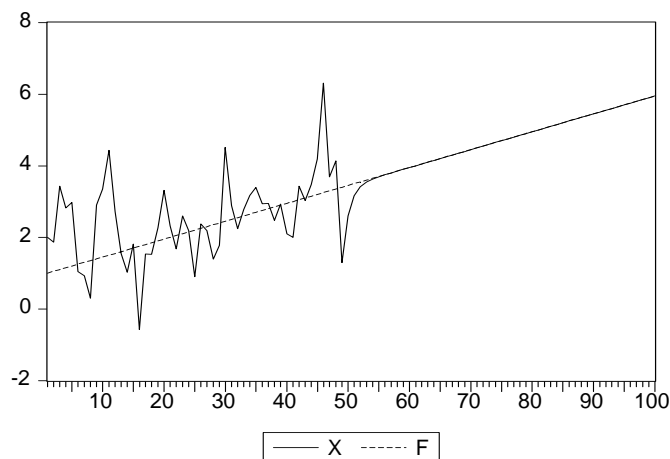
```

'- CREATION DE LA SERIE DE CHOCS-
SMPL 1 50
GENR EPS = NRND
SMPL 51 100
GENR EPS = 0
'- CREATION DE LA COMPOSANTE STATIONNAIRE Z
SMPL 1 100
GENR Z = 1
SMPL 2 100
GENR Z = 0.4*Z(-1)+EPS
'- CREATION DE LA TENDANCE F
SMPL 1 100
GENR F = 1+0.05*@TREND(1)
'- CREATION DE LA SERIE X
GENR X = F+Z

```

Ce programme débute par la simulation d'une réalisation des chocs tirés dans une loi normale  $N(0, 1)$  par l'instruction NRND sur la période de 1 à 50, puis par l'annulation des chocs pour  $t$  allant de 51 à 100. Ensuite on construit une réalisation de la composante stationnaire  $z_t$ . On commence pour cela par initialiser la série  $z_t$  à l'unité (puisque  $z_0 = 1$ ) sur l'ensemble de la période, puis on génère un  $AR(1)$  grâce à l'instruction  $Z(-1)$  qui correspond aux valeurs retardées d'une période de la série  $z_t$ . Enfin, on crée la composante tendancielle déterministe  $f(t)$  à l'aide de l'instruction @TREND( $N$ ) qui permet de générer un trend prenant la valeur 0 à la date  $N$ .

Figure 1.4: Illustration de la Propriété de Non Persistance des Chocs : Processus  $TS$



Sur le graphique (1.4) on a reporté les deux séries  $F$  et  $X$  en créant un groupe sous Eviews. On vérifie sur cette figure qu'à partir de la date d'arrêt des chocs ( $T = 50$ ), la réalisation du processus  $x_t$  converge vers la tendance déterministe. Les chocs antérieurs à la date  $T$  voient leur influence diminuer au fur et à mesure que le temps passe. On vérifie bien ici la propriété de non persistance des chocs propre au processus  $TS$ .

## 1.2 Les processus DS

Comme nous l'avons précédemment mentionné, il existe une autre forme de non stationnarité, provenant non pas de la présence d'une composante déterministe tendancielle, mais d'une source stochastique. C'est pourquoi nous allons à présent introduire la définition des processus  $DS$  pour *Difference Stationnary*.

**Definition 3** *Un processus non stationnaire*  $(x_t, t \in \mathbb{Z})$  *est un processus DS (Difference Stationnary) d'ordre*  $d$ , *où*  $d$  *désigne l'ordre d'intégration, si le processus filtré défini par*  $(1 - L)^d x_t$  *est stationnaire. On dit aussi que*  $(x_t, t \in \mathbb{Z})$  *est un processus intégré d'ordre*  $d$ , *noté*  $I(d)$ .

Ainsi, on peut définir une classe de processus stochastiques qui ne satisfont pas les conditions de la stationnarité, mais dont la différence à l'ordre  $d$  satisfait elle les propriétés de la stationnarité. Par exemple, si un processus  $z_t$  n'est pas stationnaire, on dit que ce processus est  $DS$ , intégré d'ordre un, noté  $I(1)$ , si le processus défini par la différence première  $\Delta z_t = z_t - z_{t-1}$  est quant à lui stationnaire. De la même façon, le processus  $z_t$  est  $I(2)$  si le processus défini par la différence seconde  $(1 - L)^2 z_t = (1 - L) \Delta z_t = z_t - 2z_{t-1} + z_{t-2}$  est stationnaire.

On comprend alors que la définition des processus  $DS$  repose sur *la présence de racines unitaires* dans le polynôme associé à la dynamique autorégressive du processus.

**Propriété 1.** **Un processus non stationnaire**  $(x_t, t \in \mathbb{Z})$  **est un processus DS intégré d'ordre**  $d$ , **noté**  $I(d)$ , **si le polynôme**  $\Phi(L)$  **défini en l'opérateur retard**  $L$ , **associé à sa composante autorégressive admet**  $d$  **racines unitaires :**

$$\Phi(L) x_t = z_t \quad \text{avec} \quad \Phi(L) = (1 - L)^d \tilde{\Phi}(L) \quad (1.13)$$

**où**  $z_t$  **est un processus stationnaire, et si les racines du polynôme**  $\tilde{\Phi}(L)$  **sont toutes supérieures strictement à l'unité en module.**

En effet, on pose  $\Phi(L) x_t = z_t$  avec  $\Phi(L) = (1 - L)^d \tilde{\Phi}(L)$ . Si l'on admet que les racines du polynôme  $\tilde{\Phi}(L)$  sont inférieures à l'unité en module, ce polynôme est inversible. On peut alors écrire la différence  $d^{\text{ème}}$  de  $x_t$  sous la forme d'une somme de valeurs retardées de  $z_t$  :

$$(1 - L)^d \tilde{\Phi}(L) x_t = z_t \iff (1 - L)^d x_t = \Psi(L) z_t \quad (1.14)$$

où  $\Psi(L) = \tilde{\Phi}^{-1}(L)$ . Si  $z_t$  est un processus stationnaire, la somme pondérée de ses valeurs passées  $\Psi(L)z_t$  est elle aussi stationnaire. Donc, en résumé, dès lors que le polynôme autorégressif  $\Phi(L)$  de  $x_t$  admet  $d$  racines unitaires, la quantité  $(1-L)^d x_t$  est stationnaire, donc le processus  $x_t$  est  $I(d)$ .

EXEMPLE : Considérons le processus ARMA(2,2) suivant  $\Phi(L)x_t = \Theta(L)\varepsilon_t$  avec  $\Phi(L) = (1 - 2.5L + 1.5L^2)$  et  $\Theta(L) = (1 - 0.5L)$ , et  $\varepsilon_t$  i.i.d.  $(0, \sigma_\varepsilon^2)$ . On admet que  $x_t$  est non stationnaire et l'on cherche à déterminer si  $x_t$  est un processus  $I(d)$  et quel est alors son degré d'intégration. Pour cela il suffit de déterminer le nombre de racines unitaires de  $\Phi(L)$ . Soient  $\lambda_1$  et  $\lambda_2$  les racines de  $\Phi(L) = 0$ , on a  $\lambda_1 = 1$  et  $\lambda_2 = 2/3$ . Dès lors, le processus  $x_t$  est  $I(1)$ , en effet :

$$\Phi(L) = \left(1 - \frac{1}{\lambda_1}L\right) \left(1 - \frac{1}{\lambda_2}L\right) = (1-L) \left(1 - \frac{3}{2}L\right) (1-L)^d \tilde{\Phi}(L)$$

où  $\tilde{\Phi}(L) = (1 - \frac{3}{2}L)$  admet une racine  $\lambda_2 = 2/3$  inférieure à un en module.

Dans la classe générale des processus DS, un type de processus apparaît de façon régulière, si bien que l'on lui a attribué un nom particulier : *la marche aléatoire*.

**Definition 4** Une marche aléatoire (Random Walk), ou martingale, est un processus AR(1) intégré d'ordre un, noté  $I(1)$  :

$$\Delta x_t = (1-L)x_t = c + \varepsilon_t \iff x_t = c + x_{t-1} + \varepsilon_t \quad (1.15)$$

où  $\varepsilon_t$  est un bruit blanc i.i.d.  $(0, \sigma_\varepsilon^2)$ . Si  $c = 0$ , on parle d'une marche aléatoire pure (Pure Random Walk).

Le terme de marche aléatoire provient du fait que la réalisations du processus à la date  $t$ , "part de l'endroit" où s'était arrêté  $x_{t-1}$  (la réalisation de  $x_{t-1}$ ) et va dans une direction (le choc  $\varepsilon_t$ ) totalement aléatoire. Les processus de martingale sont ainsi souvent employés en finance, où sous des propriétés particulières (marchés complets etc..) ils permettent de caractériser le cours d'un actif et le résultat selon lequel *il n'y a pas de meilleure prédicteur pour le cours de demain que le cours d'aujourd'hui*.

Une des principales propriétés des processus DS est l'hystérésis ou la persistance des chocs.

**Propriété 2.** L'influence d'une innovation  $\varepsilon_t$  à une date  $T$  sur un processus  $I(d)$

$$(1-L)^d x_t = \Theta(L)\varepsilon_t \quad (1.16)$$

est permanente. On a ainsi une propriété de persistance des chocs ou d'hystérésis.

Cela signifie que, contrairement au cas des processus  $TS$ , les chocs aléatoires  $\varepsilon_t$  conservent une influence sur le niveau de la variable  $I(d)$  et cela jusqu'à l'infini des temps. Pour bien comprendre cette propriété, nous allons considérer le cas d'une marche aléatoire avec dérive :

$$x_t = c + x_{t-1} + \varepsilon_t \quad (1.17)$$

où  $\varepsilon_t$  *i.i.d.*  $(0, \sigma_\varepsilon^2)$ . On sait que ce processus  $I(1)$  est, par définition, non stationnaire, donc on ne peut pas lui appliquer la décomposition de Wold (*cf.* chapitre précédent). Pour autant, rien ne nous empêche d'exprimer  $x_t$  sous la forme d'une moyenne mobile infinie dans les innovations passées  $\varepsilon_{t-j}$ . En effet, on a  $x_t = c + x_{t-1} + \varepsilon_t$ , en substituant  $x_{t-1}$  par son expression on obtient  $x_t = 2c + x_{t-2} + \varepsilon_{t-1} + \varepsilon_t$ . En itérant  $t$  fois cette substitution, on obtient finalement :

$$x_t = x_0 + t.c + \sum_{j=1}^t \varepsilon_j \quad (1.18)$$

On constate<sup>2</sup> à partir de *cette écriture moyenne mobile infinie que le processus  $(x_t, t \in \mathbb{Z})$  correspond à une accumulation des chocs passés et présent.* Dès lors, un choc à une date  $T$  quelconque à un impact permanent sur le niveau processus pour toutes les dates ultérieures. Supposons ainsi pour simplifier que  $c = 0$  et qu'il n'y ait qu'un seul choc  $\varepsilon_T$  à la date  $T$  ( $\varepsilon_t = 0, \forall t \neq T$ ). Le processus  $x_t$  à la date  $T$  est alors défini par  $x_T = x_0 + \varepsilon_T$ . Pour tout  $k \geq 0$ , on a  $x_{T+k} = x_0 + \varepsilon_T$ . Dès lors, lorsque  $k$  tend vers l'infini, le niveau du processus  $x_{T+k}$  ne rejoint pas la valeur initiale  $x_0$ . *L'effet du choc est donc permanent.*

Sur la figure (1.5), on a simulé deux marches aléatoires avec et sans dérive, pour un échantillon de taille 100. La première noté  $x_t$  correspond au processus  $x_t = x_{t-1} + \varepsilon_t$  et la seconde au processus  $y_t = y_{t-1} + 0.05 + \varepsilon_t$  avec  $\varepsilon_t$  *i.i.d.*  $N(0, 1)$  et  $x_1 = y_1 = 1$ . Le programme utilisé sous Eviews est le suivant :

```

!- CREATION DE LA SERIE D INNOVATIONS-
SMPL 1 1000
GENR EPS = NRND
    
```

---

<sup>2</sup>On vérifie au passage que la marche aléatoire  $x_t$  est non stationnaire puisque si l'on suppose que la condition initiale  $x_0$  est donnée et exogène, on a :

$$E(x_t) = \mu_t = x_0 + t.c \quad (1.19)$$

$$\lim_{t \rightarrow \infty} E[(x_t - \mu_t)^2] = \lim_{t \rightarrow \infty} (t\sigma_\varepsilon^2) = \infty \quad (1.20)$$

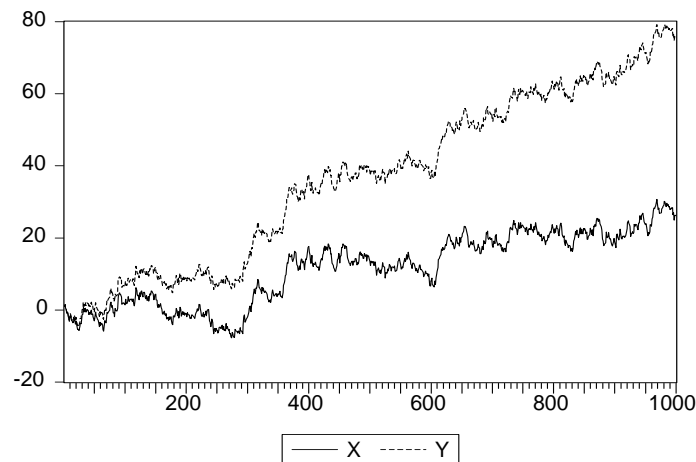
$$\gamma(h) = E[(x_t - \mu_t)(x_{t-h} - \mu_{t-h})] = \sigma_\varepsilon^2 \min(t, t-h) \quad \forall h \in \mathbb{Z} \quad (1.21)$$

Dès lors, les trois conditions de la stationnarité du second ordre sont violées. Ceci est vrai, même si la constante  $c$  est nulle puisque dans ce cas les première (équation 1.20) et troisième conditions (équation 1.21) demeurent invalides. Pour stationnariser la série, il suffit de lui appliquer le filtre  $(1-L)^d$ , avec ici  $d = 1$ . En effet,  $(1-L)x_t = \varepsilon_t$ , où  $\varepsilon_t$  est un bruit blanc, donc la série  $(1-L)x_t$  est bien stationnaire.

```
!- INITIALISATIONS DES SERIES
SMPL 1 1000
GENR X = 1
GENR Y = 1
!- CREATION DES SERIES X ET Y
SMPL 2 1000
GENR X = X(-1)+EPS
GENR Y = Y(-1)+0.05+EPS
SMPL 1 1000
```

Le programme n'appelle aucun commentaire particulier. Pour les deux réalisations obtenues sur la figure (1.5), on constate qu'a priori il n'existe aucun phénomène de "rattrapage" vers une quelconque tendance linéaire de long terme pour ces deux processus. *Mais en fait, nous verrons que leur tendance est stochastique.*

Figure 1.5: Simulations de Marches Aléatoires



### 1.3 Processus $DS$ ou $TS$ : quels enjeux ?

A ce niveau de l'exposé, on peut se demander quels sont les enjeux associés à la distinction entre les notions de non stationnarité déterministe et de non stationnarité stochastique. En effet après tout, jusqu'en maîtrise les étudiants ont déjà eu une certaine pratique de l'économétrie sans connaître la notion de stationnarité. Alors, après tout, qu'est ce que cela change que les séries soient  $TS$ ,  $DS$  etc..., pourrait se demander l'étudiant pressé de cliquer sur sa souris. Nous allons montrer que les conséquences sont doubles, sur le plan statistique et sur le plan économique.

#### 1.3.1 Conséquences statistiques de la non stationnarité

Ce que nous allons montrer tout au long de ce chapitre, c'est que si on lève l'hypothèse de stationnarité, et qu'en particulier on considère des processus de la classe  $DS$ , alors les principales méthodes d'estimation et d'inférence deviennent non fondées.

**Remarque** Les propriétés de stationnarité ou de non stationnarité des séries utilisées déterminent le type de modélisation et les propriétés asymptotiques des méthodes économétriques correspondantes.

En d'autres termes, le fait de savoir si la série statistique est une réalisation d'un processus stationnaire, non stationnaire  $DS$  ou non stationnaire  $TS$  conditionne d'une part le choix du modèle économétrique qui doit être utilisé. Mais de façon plus fondamentale et insidieuse, cela conditionne les propriétés asymptotiques des estimateurs des paramètres de ce modèle et donc par conséquent les propriétés asymptotiques des statistiques des tests usuels sur les paramètres. Si le processus est stationnaire on retrouve les propriétés standard du cours d'économétrie de base, mais si le processus est non stationnaire, et en particulier  $DS$ , on a alors des propriétés asymptotiques particulières.

Mais après tout, dira l'économètre "cliqueur", moi les propriétés asymptotiques des  $MCO$  et des statistiques de test de Student c'est pas mon problème ! Certes, mais l'ignorance de ces propriétés asymptotiques particulières peut conduire, par exemple dans le cas d'un processus  $DS$  à des erreurs de diagnostics et à des modélisations totalement fallacieuses. Prenons un exemple concret : le seuil asymptotique de significativité à 5% d'une statistique de Student d'un test de nullité sur un coefficient. Tout économètre, même "cliqueur", doit savoir que du fait de l'approximation de la loi de Student par une loi normale  $N(0, 1)$ , ce seuil est asymptotiquement égal à 1.96 dans le cas standard. Ce seuil est en particulier valide dans le cas d'une régression entre deux processus stationnaires. Or, nous allons montrer que lorsque l'on régresse deux processus  $I(1)$ , la loi asymptotique de la statistique de Student associé au test de la nullité du coefficient estimé n'est plus une loi de Student, ni une loi normale centrée réduite. Dès lors, le fameux seuil à 5% de 1.96 n'est

plus valide. Si l'économètre "cliqueur" continue d'utiliser ce seuil pour établir son diagnostic ou s'il continue d'utiliser les *pvalues* fournies par le logiciel d'économétrie sur la base d'une distribution normale, *il peut commettre et il commettra souvent des erreurs de diagnostic*. Ainsi, il acceptera à tort la significativité d'une variable ou au contraire il rejettera à tort la significativité d'une autre.

Pour bien prendre la mesure des enjeux statistiques, nous allons mener une petite expérience sous Eviews. On simule deux marches aléatoires  $x_t$  et  $y_t$  qui n'ont aucun lien entre elles :

$$x_t = x_{t-1} + \varepsilon_t$$

$$y_t = y_{t-1} + \mu_t$$

avec  $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$  et  $\mu_t \sim N(0, \sigma_\mu^2)$ . On pose  $T = 1000$ ,  $\sigma_\mu^2 = \sigma_\varepsilon^2 = 1$ . A partir de deux réalisations de ces deux processus on estime le modèle suivant par la méthode des *MCO* :

$$x_t = \beta_0 + \beta_1 y_t + \eta_t \quad (1.22)$$

De façon théorique, on sait que  $\beta_1 = 0$ , puisqu'il n'existe aucune corrélation théorique entre les deux variables. Le programme sous Eviews est le suivant :

```

!- CREATION DES SERIES D INNOVATIONS-
SMPL 1 1000
GENR EPSX = NRND
GENR EPSY = NRND
!- INITIALISATIONS DES SERIES
SMPL 1 1000
GENR X = 1
GENR Y = 1
!- CREATION DES SERIES X ET Y
SMPL 2 1000
GENR X = X(-1) + EPSX
GENR Y = Y(-1) + EPSY
SMPL 1 1000
!- REGRESSION
LS X C Y

```

Les résultats de l'estimation sont reportés sur la figure (1.6). On constate que si l'on s'en tient à la théorie standard et que la test la significativité du coefficient  $\beta_1$ , on conclut que  $\beta_1$  est différent de zéro puisque la statistique de Student associée est très largement supérieure au seuil 1.96 à 5%. La variable  $y_t$  est donc largement significative, alors que les deux marches aléatoires sont totalement indépendantes.



Figure 1.6: Régression de Deux Marches Aléatoires

Dependent Variable: X  
 Method: Least Squares  
 Date: 09/25/01 Time: 22:44  
 Sample: 1 1000  
 Included observations: 1000

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	12.23974	0.264183	46.33048	0.0000
Y	-0.542191	0.025930	-20.90939	0.0000
R-squared	0.304628	Mean dependent var	11.47154	
Adjusted R-squared	0.303931	S.D. dependent var	9.916047	
S.E. of regression	8.273033	Akaike info criterion	7.065877	
Sum squared resid	68306.19	Schwarz criterion	7.075693	
Log likelihood	-3530.939	F-statistic	437.2025	
Durbin-Watson stat	0.018627	Prob(F-statistic)	0.000000	

Cet exemple de simulations est ce que l'on appelle une *régression fallacieuse* (*Spurious Regression*). Dans la "vraie vie", on peut ainsi multiplier cet exercice à l'infini. Il suffit de considérer deux séries non stationnaires, par exemple deux séries possédant une tendance croissante relativement similaire, et de les régresser l'une sur l'autre. En appliquant les théories asymptotiques standards, vous montrerez ainsi que la production de sous vêtements féminins au Burkina Faso est une variable explicative très importante dans la détermination du cours de l'action Microsoft sur la place de New York. Il ne restera plus alors qu'à trouver une justification économique...

### 1.3.2 Conséquences économiques

Les conséquences économiques de l'introduction des processus *DS* furent toutes aussi importantes que les conséquences statistiques. *La mise évidence de la non stationnarité d'origine stochastique a tout d'abord conduit à une mise en cause générale des schémas de décomposition tendance / cycle.* Ce type de décomposition est utilisée dans de nombreux champs de l'économie appliquée (par exemple en finance dans le cadre de l'analyse chartiste etc.), mais plus particulièrement en macroéconomie. En effet, en macroéconomie appliquée, la décomposition des principales séries, comme le PIB, le taux de chômage, en une composante tendancielle et un écart conjoncturel est très souvent employée. Sur le plan théorique, elle se justifie par la relative indépendance des théories traditionnelles de la croissance par rapport aux théories des fluctuations conjoncturelles, souvent inspirés des thèses keynésiennes ou monétaristes. Jusqu'à la fin des années 80, les macroéconomistes effectuaient cette décomposition à l'aide de l'extraction d'une tendance déterministe des principales séries macroéconomiques.

Dans le cas du PIB, la tendance était alors assimilé au PIB potentiel, et les écarts à la tendance estimée correspondaient aux fluctuations conjoncturelles, c'est à dire aux cycles économiques (*GNP-*

*gap*). Or déjà à la suite de la crise économique des années 70, la rupture de rythme de croissance des économies occidentales, avait conduit à s'interroger sur cette méthode de décomposition, puisque une composante tendancielle affine du temps ne permet pas de rendre compte de cette évolution. Les plus optimistes assimilaient le ralentissement économique à l'effet transitoire d'un choc sur la composante d'écart conjoncturel. Pour d'autres, au contraire, les années 70 marquaient une rupture de tendance dans le schéma de croissance. Mais dès lors que l'on autorise la présence d'une rupture de tendance, on tend à remettre en cause le statut déterministe de la composante tendancielle. La date de rupture est elle aléatoire ? Le phénomène de rupture peut il se reproduire ? Quel aurait été la signification d'une tendance déterministe présentant des ruptures à toutes dates ? La réponse à toutes ces questions est venue avec la remise en cause de l'assimilation de la composante tendancielle à une composante déterministe et donc au choix de la méthode de décomposition.

L'extraction d'une tendance est en effet une méthode de stationnarisation propre aux processus *TS*, et ne s'applique pas aux processus *DS*. Nous allons montrer dans ce chapitre que la régression d'un processus *DS* sur une tendance déterministe peut engendrer des résultats totalement fallacieux. Chan, Hayya et Ord (1977) furent les premiers à mettre en évidence ce résultat, ils montrent en particulier que :

*"L'élimination d'une tendance linéaire d'une marche aléatoire crée artificiellement une forte autocorrélation positive des résidus dans les premiers retards"*, Chan, Hayya et Ord (1977), p. 741.

Cela signifie que si la série étudiée est *DS*, et en particulier une marche aléatoire, le fait de la régresser sur une tendance, comme le faisaient à l'époque les macroéconomistes, va conduire à une forte autocorrélation, totalement fallacieuse, des résidus, qui correspondent en fait à la composante d'écarts conjoncturels dans la décomposition tendance / cycle. Or cette autocorrélation déterminent en fait les caractéristiques cycliques des fluctuations conjoncturelles. Nelson et Kang (1981, 1984), à partir de simulations, montrent ainsi que la composante conjoncturelle présente une évolution pseudo périodique alors qu'aucun facteur n'intervenait dans les processus générateurs de données. Les cycles économiques proviendraient ils d'un artefact statistique ?

*"L'enjeu est important pour l'analyse conjoncturelle et la vision traditionnelle d'une dissociation entre cycle et croissance. A la limite, l'existence même du phénomène cyclique relèverait-il de la construction artificielle de séries en écarts à une tendance déterministe ? Slutz et Wasserfallen (1985) trouvent en effet que les cycles disparaissent dans plusieurs économies de l'OCDE lorsque les séries d'activité sont considérées comme DS et non comme TS. On mesure de ce fait l'importance que revêt la possibilité de tests spécifiques de la nature des séries macroéconomiques"*, Hénin (1989), p. 666.

La question qui se pose à présent est de savoir, si les principales séries utilisées en macroéconomie sont issues de processus  $DS$  ou de processus  $TS$ . Après tout, si la non stationnarité stochastique est un phénomène marginal, le résultat de Chan, Hayya et Ord n'aurait que peu d'importance. Ainsi, Nelson et Plosser (1982) furent les premiers à appliquer de façon systématique un ensemble de tests<sup>3</sup> de l'hypothèse de racine unitaire, c'est à dire de tests de l'hypothèse  $DS$ , à un ensemble de 14 séries macroéconomiques américaines annuelles sur des durées de 60 ans à un siècle et se terminant en 1970. Ils considèrent le PNB réel, le PNB nominal, le PNB par tête, la production industrielle, diverses séries de prix, de salaires et de rendement, la monnaie et sa vitesse de circulation ainsi que le taux de chômage. *A la seule exception du taux de chômage, les tests effectués ne permettent pas de rejeter l'hypothèse de séries  $DS$ .* Ainsi, à l'exception du taux de chômage, toutes les séries macroéconomiques américaines sont issues de processus  $DS$  et non  $TS$ . Des conclusions similaires ont été obtenues pour la plupart des pays de l'OCDE.

Les conclusions de Nelson et Plosser (1982) ont trois principales conséquences : la première c'est que la présence de racines unitaires implique que, pour une part au moins, les impulsions conjoncturelles sont constituées de chocs permanents. L'influence d'un choc affecte de façon permanente le niveau de la variable observé, c'est la propriété de persistance des chocs des processus  $DS$ . Une telle propriété peut par exemple expliquer l'absence de phénomène de "rattrapage" sur le PNB après le choc des années 70. Les séries comporte alors des tendances stochastiques. La seconde conséquence est que la plus grande part de la variabilité conjoncturelle aurait son origine dans la tendance stochastique de l'économie, c'est à dire dans les réalisations du processus de croissance lui même. La troisième implication fut notamment à l'origine du courant de pensée de nouvelle économie classique ou des cycles réels ( $RBC$  pour *Real Business Cycles*) : si l'on assimile, comme dans le keynésianisme de la synthèse, les chocs de demande à des chocs transitoires, il faut alors interpréter comme chocs d'offre les impulsions permanentes qui dominant la variabilité conjoncturelle des séries macroéconomiques américaines.

---

<sup>3</sup>Les tests de Dickey Fuller et de Dickey Fuller Augmentés utilisés par les auteurs seront présentés dans les sections 3 et 4 de ce chapitre.

## 2 Éléments de théorie asymptotique pour les processus $I(1)$

On cherche à présent à connaître les propriétés asymptotiques d'estimateurs usuels comme les MCO dans des modèles économétriques incluant des séries intégrées. Nous nous limiterons au cas de séries intégrées d'ordre 1, noté  $I(1)$ . *Ces propriétés nous permettront de non seulement comprendre les modifications des propriétés asymptotiques des MCO dans le cadre de régressions avec variables  $I(1)$ , mais aussi de construire des tests de l'hypothèse de non stationnarité.*

### 2.1 Mouvement Brownien et Théorème Central Limite Fonctionnel

Dans cette section, nous introduirons la notion de mouvement Brownien, ou processus de Wiener, puis nous présenterons le théorème central limite fonctionnel.

#### 2.1.1 Mouvement Brownien

La définition d'un mouvement Brownien est la suivante :

**Definition 5** *Un mouvement Brownien standard  $W(\cdot)$  est un processus stochastique en temps continu qui à chaque date  $t \in [0, 1]$  associe le scalaire  $W(t)$  tel que*

(i)  $W(0) = 0$

(ii) *Pour toutes dates  $0 \leq t_1 \leq t_2 \leq \dots \leq t_k \leq 1$ , les accroissements correspondants  $[W(t_2) - W(t_1)]$ ,  $[W(t_3) - W(t_2)]$ , ... et  $[W(t_k) - W(t_{k-1})]$  sont indépendants et distribués selon une loi normale  $N(0, s - t)$*

(iii) *Pour toutes réalisations,  $W(t)$  est continu en  $t$  avec une probabilité de 1.*

Un processus stochastique en temps continu  $W(\cdot)$  associe à toute date  $t \in \mathbb{R}^+$  une variable aléatoire  $W(t)$ , alors qu'un processus en temps discret n'est défini que pour des indices de temps entiers,  $t \in \mathbb{N}$ . C'est pourquoi, on distingue généralement la notation des indices pour ces deux types de processus : un indice en *subscript* pour les processus en temps discret,  $Y_t$ , et un indice entre parenthèses pour un processus en temps continu  $W(t)$ . Un processus en temps discret est ainsi représenté par une séquence dénombrable de variables aléatoires  $\{Y_t\}_{t=1}^{\infty}$ , tandis que la réalisation d'un processus en temps continu correspond à une fonction stochastique  $W(\cdot)$ ,  $t \in [0, \infty[ \rightarrow \mathbb{R}^1$ .

Le mouvement Brownien est un processus en temps continu particulier, dont les réalisations  $W(t)$  sont définies de façon continue sur  $t \in [0, 1]$ , et donc les accroissements sont indépendants et distribués selon une loi normale. On comprend ainsi, que bien que  $W(t)$  soit continu en  $t$ , cette quantité ne peut pas être différenciée en utilisant les techniques usuelles, puisque l'accroissement

entre les dates  $t$  et  $t + \Delta$  peut aller dans n'importe quel direction et prendre n'importe quelle valeur, et cela d'autant plus que  $\Delta$  est très petit.

### 2.1.2 Théorème central limite fonctionnel

On considère une Nous avons vu que lorsqu'une série est intégrée d'ordre un, l'influence des chocs est persistante. Dès lors, si l'on suppose par exemple que les réalisations des innovations d'un processus  $x_t$  sont toutes positives, le processus considéré a *un comportement explosif*, au sens où la séquence des  $x_t$  diverge lorsque  $T$  tend vers l'infini.

Soit  $x_t$  un processus  $I(1)$  observé sur les dates  $t = 1, \dots, T$ . On sait que  $x_t$  "explose" quand  $T$  tend vers l'infini. On a donc ici une différence entre l'indice du processus et la dimension  $T$  qui tend vers l'infini. *On commence donc par se ramener à un indice de temps unique tendant vers l'infini.*

**Definition 6** *Si l'on pose  $t = T \cdot \frac{t}{T} = [r \cdot T]$  où  $[.]$  désigne la partie entière, on a :*

$$x_t = x_{[r \cdot T]} \tag{2.23}$$

avec  $r = \frac{t}{T} \in [0, 1]$ . *On a ainsi "concentré" le temps et  $x_{[r \cdot T]}$  est dite série "concentrée".*

Prenons l'exemple d'un processus  $\{x_t\}_{t=0}^T$  avec  $T = 4$ . On cherche donc à changer l'indice de temps de la séquence de la façon suivante :

$$\begin{aligned} 1 = [T \cdot r] &\iff r = \frac{1}{4} & 2 = [T \cdot r] &\iff r = \frac{1}{2} \\ 3 = [T \cdot r] &\iff r = \frac{3}{4} & 4 = [T \cdot r] &\iff r = 1 \end{aligned}$$

On peut donc changer l'indice de temps de la séquence  $\{x_0, x_1, x_2, x_3, x_4\}$  de la façon suivante  $\{x_{[T \cdot 0]}, x_{[T \cdot \frac{1}{4}]}, x_{[T \cdot \frac{1}{2}]}, x_{[T \cdot \frac{3}{4}]}, x_{[T \cdot 1]}\}$  ce qui revient à poser  $\{x_{[T \cdot r]}\}$  pour  $r = \{0, \frac{1}{4}, \frac{1}{2}, \frac{3}{4}, 1\}$  et  $T = 4$ . Mais attention, la série concentrée  $x_{[r \cdot T]}$  est définie pour un indice  $r$  continu sur  $[0, 1]$ . Ainsi, si  $r < 1/4$ , alors  $[T \cdot r] = 0$ , donc on a  $x_{[T \cdot r]} = x_0$ . Pour<sup>4</sup>  $r \in [1/4, 1/2[$  alors  $[T \cdot r] = 1$ , donc  $x_{[T \cdot r]} = x_1$ . La définition du processus concentré  $x_{[r \cdot T]}$  pour  $r \in [0, 1]$  est donc la suivante :

$$x_{[r \cdot T]} = \begin{cases} x_0 & r < \frac{1}{4} \\ x_1 & \frac{1}{4} \leq r < \frac{1}{2} \\ x_2 & \frac{1}{2} \leq r < \frac{3}{4} \\ x_3 & \frac{3}{4} \leq r < 1 \\ x_4 & r = 1 \end{cases} \tag{2.24}$$

De plus, lorsque  $T$  tend vers l'infini l'indice de la série concentrée  $x_{[T \cdot r]}$ ,  $r \in [0, 1]$  converge lui aussi vers l'infini.

---

<sup>4</sup>Par exemple  $r = 0.3 \in [1/4, 1/2[$ , on a  $[T \cdot r] = [4 \cdot 0.3] = [1.2] = 1$ .

En second lieu, il faut contrôler la vitesse d'explosion de la série. On sait que si  $x_t$  est  $I(1)$ , ce processus "explose" en  $t^{1/2}$  quand  $T$  tend vers l'infini. Mais de façon générale on note  $\alpha$  la vitesse d'explosion d'un processus qui explose en  $t^\alpha$ . On cherche alors à contrôler cette vitesse de convergence de la façon suivante :

**Definition 7** *Si le processus  $x_t$  explose en  $t^\alpha$ ,  $\alpha > 0$ , on note :*

$$X_T(r) = \frac{x_{[Tr]}}{T^\alpha} \quad r \in [0, 1] \quad (2.25)$$

On cherche alors à caractériser  $X_T(r)$  (qui est un processus indexé par un ensemble continu sur  $r$ ), lorsque  $T$  tend vers l'infini, pour tout  $r \in [0, 1]$  en choisissant  $\alpha$  de telle sorte que  $X_T(r)$  n'ait plus de comportement explosif en  $T$ . Pour cela nous allons à présent considérer l'exemple de la marche aléatoire pure.

On considère un processus  $x_t$  représenté par une marche aléatoire pure :

$$x_t = x_{t-1} + \varepsilon_t \quad (2.26)$$

avec  $\varepsilon_t$  i.i.d.  $(0, \sigma_\varepsilon^2)$ . On pose  $x_0 = 0$ , dès lors on a  $x_t = \sum_{j=1}^t \varepsilon_j$ . On sait que le processus  $x_t$  explose à la vitesse  $T^{\frac{1}{2}}$ . On définit donc la série suivante :

$$X_T(r) = T^{-\frac{1}{2}} x_{[Tr]} \quad r \in [0, 1] \quad (2.27)$$

Or, dans ce cas on montre que :

$$X_T(r) = T^{-\frac{1}{2}} \sum_{j=1}^{[Tr]} \varepsilon_j \quad r \in [0, 1] \quad (2.28)$$

En fait, le processus  $X_T(r)$  correspond à un processus en marche d'escalier. Pour tout  $r < \frac{1}{T}$ , on a  $t = [r.T] = 0$  et donc on a  $x_{[Tr]} = x_0$

**Theorem 8 (Théorème Central Limite Fonctionnel TCLF)** *Soit  $v_t$  un vecteur de dimension  $n$  de variables aléatoires i.i.d.  $(0, \Sigma)$ , avec  $\Sigma = PP'$ . On pose,  $\forall t = 1, \dots, T$*

$$u_t = \Psi(L) v_t = \sum_{s=0}^{\infty} \Psi^s v_{t-s}$$

où les éléments  $\Psi_{ij}^s$  du polynôme matriciel  $\Psi(L)$  vérifient  $\sum_{s=0}^{\infty} s \left| \Psi_{ij}^s \right| < \infty \forall i, \forall j = 1, \dots, n$ . La séquence de fonctions stochastiques vectorielles  $\left\{ \sqrt{T} X_T^*(\cdot) \right\}_{T=1}^{\infty}$  définies par

$$X_T^*(r) = \frac{1}{T} \sum_{t=1}^{[Tr]} u_t \quad \forall r \in [0, 1]$$

a une distribution asymptotique décrite par un mouvement Brownien vectoriel  $(n, n)$  standard, noté  $W^*(\cdot)$ .

$$\sqrt{T}X_T^*(\cdot) \xrightarrow[T \rightarrow \infty]{\mathcal{L}} \Psi(1)PW^*(\cdot)$$

où  $\Psi(1)PW^*(r)$  est distribué suivant une loi normale  $N(0, r[\Psi(1)\Sigma\Psi(1)'])$ .

Dans le cas  $r = 1$ , on retrouve le TCL standard.

**Theorem 9 (Continuous Mapping Theorem)** Soit  $S_T(\cdot)$  une fonction stochastique telle que  $S_T(\cdot) \xrightarrow{\mathcal{L}} S(\cdot)$  et soit  $g(\cdot)$  une fonctionnelle continue, alors  $g[S_T(\cdot)] \xrightarrow{\mathcal{L}} g[S(\cdot)]$ .

## 2.2 Application aux processus $I(1)$

La première utilisation du théorème central limite fonctionnel pour établir les distributions asymptotiques de statistiques construites à partir de processus  $I(1)$  est due à Phillips (1986, 1978). On se limitera ici à l'étude du cas simple d'une marche aléatoire. Pour l'étude à des processus dont les innovations sont autocorrélées, le lecteur pourra se référer à Phillips (1986, 1978). Ces distributions asymptotiques nous seront particulièrement utiles pour la construction des tests de non stationnarité et pour bien comprendre les enjeux statistiques de la non stationnarité statistique.

L'illustration la plus simple des résultats de Phillips conduit à étudier une marche aléatoire pure :

$$x_t = x_{t-1} + \varepsilon_t \quad (2.29)$$

avec  $\varepsilon_t$  *i.i.d.*  $(0, \sigma_\varepsilon^2)$ . On suppose pour simplifier que la condition initiale sur le processus (2.29) est déterministe et nulle :  $x_0 = 0$ . Nous avons vu que dans ce cas là le processus en temps discret  $x_t$  correspond à l'accumulation des chocs passés entre la date 1 et la date  $t$  :

$$x_t = \varepsilon_1 + \varepsilon_2 + \dots + \varepsilon_t \quad (2.30)$$

A partir de la définition du processus  $x_t$  on peut définir la fonction continue stochastique  $X_T(r) = x_{[rT]}/T$  en fonction d'un indice de temps continu  $r \in [0, 1]$  :

$$X_T(r) = \begin{cases} \frac{1}{T}x_0 = 0 & 0 \leq r < \frac{1}{T} \\ \frac{1}{T}x_1 = \frac{\varepsilon_1}{T} & \frac{1}{T} \leq r < \frac{2}{T} \\ \frac{1}{T}x_2 = \frac{1}{T}(\varepsilon_1 + \varepsilon_2) & \frac{2}{T} \leq r < \frac{3}{T} \\ \dots & \dots \\ \frac{1}{T}x_T = \frac{1}{T}(\varepsilon_1 + \dots + \varepsilon_T) & r = 1 \end{cases} \quad (2.31)$$

La fonction stochastique  $X_T(r)$ , conditionnellement aux réalisations  $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_T$  peut ainsi être représentée sous la forme d'une fonction en "marches d'escalier", comme le montre l'illustration (2.7). Pour toutes les valeurs de  $r$  comprises entre 0 et  $1/T$ , valeur exclue, les réalisations de  $TX_T(r)$  correspondent à la condition initiale  $x_0$ . Entre  $1/T$  et  $2/T$ , valeur exclue, les réalisations de  $TX_T(r)$  correspondent à la réalisation  $\varepsilon_1$ , et ainsi de suite. Pour les valeurs de  $r$  comprises entre  $(T-1)/T$  et  $T$ , réalisations de  $TX_T(r)$  correspondent à la somme des réalisations de  $\varepsilon_1$  à  $\varepsilon_{T-1}$ .

Figure 2.7: Illustration de la Forme des Réalisations  $X_T(r)$

Bien entendu, l'aire définie par cette fonction en marches d'escalier correspond à la somme de l'aire de  $T$  rectangles. Le  $i^{\text{ème}}$  rectangle a une largeur égale à  $1/T$  et une longueur égale à  $x_{i-1}/T$ , et donc son aire est égale à  $x_{i-1}/T^2$ . Dès lors l'intégrale de la fonction stochastique  $X_T(r)$  est équivalente à la somme :

$$\int_0^1 X_T(r) dr = \frac{x_0}{T^2} + \frac{x_1}{T^2} + \frac{x_2}{T^2} + \dots + \frac{x_{T-1}}{T^2} \quad (2.32)$$

En multipliant les deux membres de cette égalité par  $\sqrt{T}$ , on obtient :

$$\sqrt{T} \int_0^1 X_T(r) dr = T^{-\frac{3}{2}} \sum_{t=1}^T x_{t-1} \quad (2.33)$$

Or, d'après le théorème central limite fonctionnel on sait que :

$$\sqrt{T} X_T(\cdot) \xrightarrow[T \rightarrow \infty]{\mathcal{L}} \sigma_\varepsilon W(\cdot) \quad (2.34)$$

On définit à présent une fonctionnelle continue  $g(\cdot)$ , telle que  $g\left[\sqrt{T} X_T(r)\right] = \int_0^1 \sqrt{T} X_T(r) dr = \sqrt{T} \int_0^1 X_T(r) dr$ , par application du *continuous mapping theorem* on obtient :



$$\sqrt{T} \int_0^1 X_T(r) dr \xrightarrow[T \rightarrow \infty]{\mathcal{L}} \sigma_\varepsilon \int_0^1 W(r) dr \quad (2.35)$$

Dès lors, on peut établir la distribution asymptotique de la somme de marches aléatoires :

$$T^{-\frac{3}{2}} \sum_{t=1}^T x_{t-1} \xrightarrow[T \rightarrow \infty]{\mathcal{L}} \sigma_\varepsilon \int_0^1 W(r) dr \quad (2.36)$$

Ce résultat implique que pour une marche aléatoire sans dérive, la somme  $T^{-\frac{3}{2}} \sum_{t=1}^T x_{t-1}$  converge vers une variable aléatoire de distribution normale  $N(0, \sigma_\varepsilon^2/3)$  dont la distribution peut être exprimée comme une intégrale de réalisations de mouvements Browniens de variance  $\sigma_\varepsilon^2$ . Il est intéressant de rappeler ici que pour une marche aléatoire sans dérive, la moyenne empirique  $T \sum_{t=1}^T x_t$  diverge.

De la même façon, en utilisant le théorème central limite fonctionnel ainsi que le *continuous mapping theorem*, on peut dériver la distribution asymptotique de la plupart des moments empiriques d'une marche aléatoire. Ces résultats sont résumés dans la proposition (10).

**Proposition 10** *On considère un processus  $x_t$  satisfaisant une représentation AR(1) non stationnaire I(1) telle que :*

$$x_t = x_{t-1} + \varepsilon_t \quad (2.37)$$

*avec  $x_0 = 0$  et où  $\varepsilon_t$  i.i.d.  $(0, \sigma_\varepsilon^2)$ . Les distributions asymptotiques des principaux moments empiriques de  $x_t$  sont alors les suivantes :*

$$T^{-1/2} \sum_{t=1}^T \varepsilon_t \xrightarrow[T \rightarrow \infty]{\mathcal{L}} \sigma_\varepsilon W(1) \quad (2.38)$$

$$T^{-1} \sum_{t=1}^T x_{t-1} \varepsilon_t \xrightarrow[T \rightarrow \infty]{\mathcal{L}} \frac{1}{2} \sigma_\varepsilon^2 \{ [W(1)]^2 - 1 \} \quad (2.39)$$

$$T^{-3/2} \sum_{t=1}^T t \varepsilon_t \xrightarrow[T \rightarrow \infty]{\mathcal{L}} \sigma_\varepsilon W(1) - \sigma_\varepsilon \int_0^1 W(r) dr \quad (2.40)$$

$$T^{-3/2} \sum_{t=1}^T x_{t-1} \xrightarrow[T \rightarrow \infty]{\mathcal{L}} \sigma_\varepsilon \int_0^1 W(r) dr \quad (2.41)$$

$$T^{-2} \sum_{t=1}^T x_{t-1}^2 \xrightarrow[T \rightarrow \infty]{\mathcal{L}} \sigma_\varepsilon^2 \int_0^1 [W(r)]^2 dr \quad (2.42)$$

$$T^{-5/2} \sum_{t=1}^T t x_{t-1} \xrightarrow[T \rightarrow \infty]{\mathcal{L}} \sigma_\varepsilon \int_0^1 r W(r) dr \quad (2.43)$$

$$T^{-3} \sum_{t=1}^T t x_{t-1}^2 \xrightarrow[T \rightarrow \infty]{\mathcal{L}} \sigma_\varepsilon^2 \int_0^1 r [W(r)]^2 dr \quad (2.44)$$

$$T^{-(v+1)} \sum_{t=1}^T t^v \xrightarrow[T \rightarrow \infty]{} 1/(v+1) \quad v \in \mathbb{N} \quad (2.45)$$

où  $W(\cdot)$  désigne un mouvement Brownien standard.

Nous ne démontrerons pas l'ensemble des ces convergences. Nous nous limiterons aux résultats (2.39) et (2.42). Commençons par le résultat (2.42). Pour cela, on définit un nouveau processus en temps continu  $S_T(r)$  tel que :

$$S_T(r) \equiv T [X_T(r)]^2 \quad \forall r \in [0, 1] \quad (2.46)$$

Ce nouveau processus  $S_T(r)$  correspond en fait à la somme partiel des carrés des innovations  $\varepsilon_t$  puisque :

$$S_T(r) = \begin{cases} \frac{1}{T} x_0^2 = 0 & 0 \leq r < \frac{1}{T} \\ \frac{1}{T} x_1^2 = \frac{1}{T} \varepsilon_1^2 & \frac{1}{T} \leq r < \frac{2}{T} \\ \frac{1}{T} x_2^2 = \frac{1}{T} (\varepsilon_1 + \varepsilon_2)^2 & \frac{2}{T} \leq r < \frac{3}{T} \\ \dots & \dots \\ \frac{1}{T} x_T^2 = \frac{1}{T} (\varepsilon_1 + \dots + \varepsilon_T)^2 & r = 1 \end{cases} \quad (2.47)$$

En faisant le même raisonnement que précédemment, on montre que :

$$T^{-2} \sum_{t=1}^T x_{t-1}^2 = \frac{x_0^2}{T^2} + \frac{x_1^2}{T^2} + \frac{x_2^2}{T^2} + \dots + \frac{x_{T-1}^2}{T^2} = \int_0^1 S_T(r) dr \quad (2.48)$$

D'après le théorème central limite fonctionnel on sait que :

$$\sqrt{T} X_T(\cdot) \xrightarrow[T \rightarrow \infty]{\mathcal{L}} \sigma_\varepsilon W(\cdot)$$

On définit à présent une fonctionnelle continue  $g(\cdot)$ , telle que à toute réalisation  $S_T(r) = T [X_T(r)]^2 = [\sqrt{T} X_T(r)]^2$  on associe  $g[S_T(r)] = \int_0^1 S_T(r) dr = \int_0^1 [\sqrt{T} X_T(r)]^2 dr$ . Par application du *continuous mapping theorem* on obtient :

$$\int_0^1 S_T(r) dr = \int_0^1 [\sqrt{T} X_T(r)]^2 dr \xrightarrow[T \rightarrow \infty]{\mathcal{L}} g[\sigma_\varepsilon W(r)] \equiv \int_0^1 [\sigma_\varepsilon W(r)]^2 dr$$

D'après ce résultat, en reprenant l'égalité (2.48), on démontre finalement le résultat (2.42) :

$$T^{-2} \sum_{t=1}^T x_{t-1}^2 = \int_0^1 S_T(r) dr \xrightarrow[T \rightarrow \infty]{\mathcal{L}} \sigma_\varepsilon^2 \int_0^1 [W(r)]^2 dr \quad (2.49)$$

Démontrons à présent le résultat (2.39) de la proposition (10) selon lequel :

$$T^{-1} \sum_{t=1}^T x_{t-1} \varepsilon_t \xrightarrow[T \rightarrow \infty]{\mathcal{L}} \frac{1}{2} \sigma_\varepsilon^2 \{ [W(1)]^2 - 1 \} \quad (2.50)$$

Pour cela il convient tout d'abord de transformer le moment empirique  $\sum_{t=1}^T x_{t-1} \varepsilon_t$ . Pour cela, on utilise le résultat suivant :

$$\left[ \sum_{t=1}^T \varepsilon_t \right]^2 = \sum_{t=1}^T \varepsilon_t^2 + 2 \sum_{t=1}^T \left( \sum_{j=1}^{t-1} \varepsilon_j \right) \varepsilon_t \quad (2.51)$$

Etant la définition de la marche aléatoire  $x_t$ , on sait que  $x_t = \sum_{s=1}^t \varepsilon_s$  puisque  $x_0 = 0$ . Dès lors cette égalité peut se réécrire sous la forme suivante :

$$x_T^2 = \sum_{t=1}^T \varepsilon_t^2 + 2 \sum_{t=1}^T x_{t-1} \varepsilon_t$$

On en déduit alors une nouvelle expression du moment empirique qui nous intéresse, à savoir  $\sum_{t=1}^T x_{t-1} \varepsilon_t$  :

$$\left( \frac{1}{T} \right) \sum_{t=1}^T x_{t-1} \varepsilon_t = \left( \frac{1}{2} \right) \left( \frac{1}{T} \right) x_T^2 - \left( \frac{1}{2} \right) \left( \frac{1}{T} \right) \sum_{t=1}^T \varepsilon_t^2 \quad (2.52)$$

Etudions la convergence de ces différents éléments. Commençons par la quantité  $T^{-1} \sum_{t=1}^T \varepsilon_t^2$ . En utilisant la loi des grands nombres (*cf.* chapitre 1), on sait que si  $\varepsilon_t$  est *i.i.d.*  $(0, \sigma_\varepsilon^2)$  alors cette quantité, qui est tout simplement le moment empirique d'ordre deux des  $\varepsilon_t$ , converge en probabilité vers le moment théorique d'ordre deux, à savoir la variance  $\sigma_\varepsilon^2$  :

$$\left( \frac{1}{T} \right) \sum_{t=1}^T \varepsilon_t^2 \xrightarrow[T \rightarrow \infty]{p} \sigma_\varepsilon^2 \quad (2.53)$$

Reste le premier éléments de l'équation (2.52). En reprenant la définition du processus en temps continu  $S_T(\cdot)$ , on montre immédiatement que  $T^{-1} x_T^2 = S_T(1) \equiv \left[ \sqrt{T} X_T(1) \right]^2$ . D'après le théorème central limite fonctionnel, on sait que  $\sqrt{T} X_T(1)$  converge en loi vers  $\sigma_\varepsilon W(1)$ . Dès lors, on définit une fonctionnelle  $g(\cdot)$  qui à toute réalisation  $\sqrt{T} X_T(r)$  associe  $g \left[ \sqrt{T} X_T(r) \right] = \left[ \sqrt{T} X_T(r) \right]^2 = S_T(r)$ . Par application du *continuous mapping theorem* on montre alors que  $S_T(1)$  converge en loi vers  $\sigma_\varepsilon^2 [W(1)]^2$ . Finalement, on montre ainsi que :

$$\left( \frac{1}{T} \right) x_T^2 = S_T(1) \xrightarrow[T \rightarrow \infty]{\mathcal{L}} g[S_T(1)] \equiv \sigma_\varepsilon^2 [W(1)]^2 \quad (2.54)$$

En reprenant les résultats (2.53) et (2.54), ainsi que l'expression (2.52) du moment empirique on retrouve le résultat (2.39) de la proposition (10) :

$$\begin{aligned} \left( \frac{1}{T} \right) \sum_{t=1}^T x_{t-1} \varepsilon_t &= \left( \frac{1}{2} \right) \left( \frac{1}{T} \right) x_T^2 - \left( \frac{1}{2} \right) \left( \frac{1}{T} \right) \sum_{t=1}^T \varepsilon_t^2 \\ &\xrightarrow[T \rightarrow \infty]{\mathcal{L}} \left( \frac{\sigma_\varepsilon^2}{2} \right) [W(1)]^2 - \left( \frac{\sigma_\varepsilon^2}{2} \right) \end{aligned} \quad (2.55)$$

Rappelons que la réalisation  $W(1)$ , c'est à dire la réalisation d'un mouvement Brownien à la date 1, a une distribution  $N(0, 1)$ . Cela signifie que la variable aléatoire  $[W(1)]^2$  a une distribution  $\chi^2(1)$ .

L'étude de ces distributions asymptotiques nous permet en particulier de définir les vitesses de convergence des différents moments empiriques de la marche aléatoire. Cette vitesse est représenté par la puissance  $\alpha$  de  $T$  pour laquelle un processus  $T^{-\alpha}z_t$  converge en probabilité vers une valeur finie.

**Definition 11** *Une séquence de variable aléatoire  $\{z_t\}_{t=1}^{\infty}$  est dite  $Op(T^\alpha)$ , si pour tout  $\varepsilon > 0$ , il existe une valeur finie  $M$  telle que :*

$$P \left\{ \left| \frac{z_T}{T^\alpha} \right| > M \right\} < \varepsilon \quad (2.56)$$

En appliquant cette définition aux résultats de la proposition (10), on détermine les vitesses de convergence des différents moments empiriques :

**Proposition 12** *Les vitesses de convergence des moments empiriques d'une marche aléatoire sont les suivantes :*

$$\begin{aligned} \sum_{t=1}^T \varepsilon_t &= Op \left( T^{\frac{1}{2}} \right) & \sum_{t=1}^T x_{t-1} \varepsilon_t &= Op(T) \\ \sum_{t=1}^T t \varepsilon_t &= Op \left( T^{\frac{3}{2}} \right) & \sum_{t=1}^T x_{t-1} &= Op \left( T^{\frac{3}{2}} \right) \\ \sum_{t=1}^T x_{t-1}^2 &= Op(T^2) & \sum_{t=1}^T t x_{t-1} &= Op \left( T^{\frac{5}{2}} \right) \\ \sum_{t=1}^T t x_{t-1}^2 &\xrightarrow[T \rightarrow \infty]{\mathcal{L}} Op(T^3) \end{aligned}$$

L'étudiant intéressé devra outre la proposition (10), étudier le cas où les innovations  $\varepsilon_t$  sont autocorrélés et le cas vectoriel, où  $x_t$  désigne un vecteur d'observation de  $n$  variables.

### 2.3 Les régressions fallacieuses

Maintenant que nous avons établi les distributions asymptotique d'un certain nombre de moments empiriques associés à une marche aléatoire, on peut s'intéresser aux propriétés asymptotiques des estimateurs des MCO et des principales statistiques de tests usuelles lorsque le modèle comprend des variables  $I(1)$ . Nous allons en particulier étudier le cas des régressions fallacieuses (*Spurious Regressions*). De façon générale, on a le résultat suivant :

**Definition 13** *On considère une régression de la forme  $y_t = \beta x_t + \mu_t$  où  $x_t$  et  $y_t$  désigne deux variables  $I(1)$ . On suppose qu'il n'existe aucune valeur de  $\beta$  telle que le résidu  $\mu_t = y_t - \beta x_t$  soit  $I(0)$ . Alors l'estimateur des MCO, appliqué à ce modèle, conduit à une phénomène de régression fallacieuse.*

Ce phénomène a été mis en évidence à partir d'expériences de Monte Carlo par Granger et Newbold (1974) et fut plus tard expliqué théoriquement par Phillips (1986). Que ce passe-t-il dans le cas des régressions fallacieuses ? Nous allons montrer que bien que les estimateurs des *MCO* soient convergents, la plupart des statistiques de tests usuelles n'ont plus une distribution standard. Dès lors, on comprend les enjeux statistiques associés à la non stationnarité : si l'on applique les théories usuelles dans des configurations de régressions fallacieuses, on pourra montrer sans aucun problème que la production de sous vêtements féminins au Burkina Faso est une variable explicative très importante dans la détermination du cours de l'action Microsoft sur la place de New York.... Nous allons tout d'abord étudier ce phénomène sur le plan théorique avant de mener des expériences de Monte Carlo.

### 2.3.1 Distributions asymptotiques dans le cas des régressions fallacieuses

Montrons à présent comment, dans le cas d'une régression fallacieuse, les propriétés asymptotiques des estimateurs et des statistiques de tests usuelles sont affectées. Pour cela nous allons considérer un petit modèle extrêmement simple.

**Hypothèses ( $H_1$ )** On considère deux marches aléatoires pures sans aucun lien :

$$x_t = x_{t-1} + \varepsilon_t \quad (2.57)$$

$$y_t = y_{t-1} + \mu_t \quad (2.58)$$

avec  $x_0 = y_0 = 0$ , et  $\varepsilon_t$  *i.i.d.*  $(0, \sigma_\varepsilon^2)$ ,  $\mu_t$  *i.i.d.*  $(0, \sigma_\mu^2)$ . On suppose que les innovations des processus  $x_t$  et  $y_t$  sont totalement indépendantes :  $E(\varepsilon_t \mu_s) = 0, \forall (s, t)$ .

Supposons qu'un économètre étourdi régresse la variable  $x_t$  sur la variable  $y_t$ , sans avoir au préalable diagnostiquer la non stationnarité de ces deux processus et l'absence de relation stable entre les deux séries. Notre économètre va donc effectuer la régression :

$$y_t = \beta_0 + \beta_1 x_t + \eta_t \quad (2.59)$$

Nous sommes ici dans une configuration de régression fallacieuse. Etudions le comportement asymptotique des estimateurs des *MCO*,  $\hat{\beta}_0$  et  $\hat{\beta}_1$  ainsi que le comportement asymptotique de la statistique de Student associée au test  $H_0 : \beta_1 = 0$ . Rappelons au passage, que puisque les deux variables  $x_t$  et  $y_t$  n'ont aucun lien, on a  $\beta_1 = \beta_0 = 0$ .

Commençons par étudier le comportement des estimateurs  $\hat{\beta}_0$  et  $\hat{\beta}_1$ . D'après la définition des estimateurs de *MCO* du modèle (2.59), on a :

$$\begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix} = \begin{pmatrix} T & \sum_{t=1}^T x_t \\ \sum_{t=1}^T x_t & \sum_{t=1}^T x_t^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum_{t=1}^T y_t \\ \sum_{t=1}^T x_t y_t \end{pmatrix} \quad (2.60)$$

Essayons à présent d'étudier la vitesse de convergence des différents éléments des estimateurs  $\hat{\beta}_0$  et  $\hat{\beta}_1$ . L'élément  $T$  diverge à la vitesse  $T$ , il est  $Op(T)$ . Concernant l'élément  $\sum_{t=1}^T x_t$ , on a :

$$\sum_{t=1}^T x_t = \sum_{t=1}^T x_{t-1} + x_T = \sum_{t=1}^T x_{t-1} + \sum_{t=1}^T \varepsilon_t \quad (2.61)$$

Nous avons vu dans la section précédent que  $\sum_{t=1}^T \varepsilon_t = Op\left(T^{\frac{1}{2}}\right)$  et que  $\sum_{t=1}^T x_{t-1} = Op\left(T^{\frac{3}{2}}\right)$ . Il convient donc de déflater  $\sum_{t=1}^T x_t$  par  $T^{3/2}$  pour obtenir une distribution non divergente :

$$\begin{aligned} T^{-3/2} \sum_{t=1}^T x_t &= T^{-3/2} \sum_{t=1}^T x_{t-1} + T^{-1} \left( T^{-1/2} \sum_{t=1}^T \varepsilon_t \right) \\ &= T^{-3/2} \sum_{t=1}^T x_{t-1} + T^{-1} \cdot Op\left(T^{\frac{1}{2}}\right) \end{aligned}$$

Cela signifie que la quantité à droite de cette expression converge en probabilité vers 0, lorsque  $T$  tend vers l'infini :

$$T^{-1/2} \sum_{t=1}^T \varepsilon_t \xrightarrow[T \rightarrow \infty]{\mathcal{L}} \sigma_\varepsilon W(1) \iff T^{-1} \left( T^{-1/2} \sum_{t=1}^T \varepsilon_t \right) \xrightarrow[T \rightarrow \infty]{\mathcal{P}} 0$$

Ainsi la quantité  $T^{-3/2} \sum_{t=1}^T x_t$  converge vers la même limite que  $T^{-3/2} \sum_{t=1}^T x_{t-1}$ . Dès lors, en utilisant la proposition (10), on montre que :

$$T^{-3/2} \sum_{t=1}^T x_t \xrightarrow[T \rightarrow \infty]{\mathcal{L}} \sigma_\varepsilon \int_0^1 W_1(r) dr \quad (2.62)$$

De la même façon, on montre que :

$$T^{-3/2} \sum_{t=1}^T y_t \xrightarrow[T \rightarrow \infty]{\mathcal{L}} \sigma_\mu \int_0^1 W_2(r) dr \quad (2.63)$$

où  $W_1(\cdot)$  et  $W_2(\cdot)$  sont deux mouvement Browniens standards indépendants.

Concernant l'élément  $\sum_{t=1}^T x_t^2$  de la définition des MCO, on sait que :

$$\sum_{t=1}^T x_t^2 = \sum_{t=1}^T x_{t-1}^2 + x_T^2 = \sum_{t=1}^T x_{t-1}^2 + \left( \sum_{t=1}^T \varepsilon_t \right)^2 \quad (2.64)$$

Par un raisonnement analogue au précédent, sachant que  $\sum_{t=1}^T x_{t-1}^2 = Op(T^2)$ , il convient de déflater  $\sum_{t=1}^T x_t^2$  par  $T^2$  pour obtenir une distribution non divergente :

$$\begin{aligned} T^{-2} \sum_{t=1}^T x_t^2 &= T^{-2} \sum_{t=1}^T x_{t-1}^2 + T^{-1} \left( T^{-1/2} \sum_{t=1}^T \varepsilon_t \right)^2 \\ &= T^{-2} \sum_{t=1}^T x_{t-1}^2 + T^{-1} \cdot \left[ Op\left(T^{\frac{1}{2}}\right) \right]^2 \end{aligned}$$

De la même façon, on montre que la quantité  $T^{-2} \sum_{t=1}^T x_t^2$  converge vers la même limite que  $T^{-2} \sum_{t=1}^T x_{t-1}^2$ . Dès lors, en utilisant la proposition (10), on montre que :

$$T^{-2} \sum_{t=1}^T x_{t-1}^2 \xrightarrow[T \rightarrow \infty]{\mathcal{L}} \sigma_\varepsilon^2 \int_0^1 [W_1(r)]^2 dr \quad (2.65)$$

où  $W_1(\cdot)$  désigne le mouvement Brownien précédemment défini.

Il ne reste plus enfin qu'à étudier la convergence de l'élément  $\sum_{t=1}^T x_t y_t$ . On admettra le résultat suivant :

$$T^{-2} \sum_{t=1}^T x_t y_t \xrightarrow[T \rightarrow \infty]{\mathcal{L}} \sigma_\varepsilon \sigma_\mu \int_0^1 W_1(r) W_2(r) dr \quad (2.66)$$

où  $W_1(\cdot)$  et  $W_2(\cdot)$  sont les deux mouvements Browniens précédemment définis.

Nous avons à présent l'ensemble des éléments nécessaires à la détermination de la distribution asymptotique des estimateurs  $\hat{\beta}_0$  et  $\hat{\beta}_1$ . Il suffit pour cela d'équilibrer l'équation (2.60) dans les termes en  $T$ , après avoir contrôlé la vitesse de convergence de chaque élément<sup>5</sup>. On obtient alors la distribution asymptotique des estimateurs  $\hat{\beta}_0$  et  $\hat{\beta}_1$ .

**Proposition 14** *Sous les hypothèses (2.3.1), les estimateurs des MCO ,  $\hat{\beta}_0$  et  $\hat{\beta}_1$ , obtenus dans le modèle :*

$$y_t = \beta_0 + \beta_1 x_t + \eta_t \quad (2.67)$$

*ont pour distribution asymptotique :*

$$\begin{aligned} \begin{pmatrix} T^{-\frac{1}{2}} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix} &= \begin{pmatrix} T^{-1} T & T^{-\frac{3}{2}} \sum_{t=1}^T x_t \\ T^{-\frac{3}{2}} \sum_{t=1}^T x_t & T^{-2} \sum_{t=1}^T x_t^2 \end{pmatrix}^{-1} \begin{pmatrix} T^{-\frac{3}{2}} \sum_{t=1}^T y_t \\ T^{-2} \sum_{t=1}^T x_t y_t \end{pmatrix} \\ &\xrightarrow[T \rightarrow \infty]{\mathcal{L}} \begin{pmatrix} 1 & \sigma_\varepsilon \int_0^1 W_1(r) dr \\ \sigma_\varepsilon \int_0^1 W_1(r) dr & \sigma_\varepsilon^2 \int_0^1 [W_1(r)]^2 dr \end{pmatrix}^{-1} \begin{pmatrix} \sigma_\mu \int_0^1 W_2(r) dr \\ \sigma_\varepsilon \sigma_\mu \int_0^1 W_1(r) W_2(r) dr \end{pmatrix} \end{aligned} \quad (2.68)$$

En explicitant les termes de la proposition (14), on montre en particulier que :

$$\hat{\beta}_1 \xrightarrow[T \rightarrow \infty]{\mathcal{L}} h_1 = \left( \frac{\sigma_\mu}{\sigma_\varepsilon} \right) \left( \frac{\int_0^1 W_1(r) W_2(r) dr - \int_0^1 W_1(r) dr \int_0^1 W_2(r) dr}{\int_0^1 [W_1(r)]^2 dr - \left[ \int_0^1 W_1(r) dr \right]^2} \right) \quad (2.69)$$

<sup>5</sup>En effet,  $\forall T \in \mathbb{R}^{+*}$ , et  $\forall (a, b, c, d, e, f) \in \mathbb{R}^6$  tels que  $ad \neq c^2$ , on a :

$$\begin{bmatrix} T^{-1}a & T^{-\frac{3}{2}}c \\ T^{-\frac{3}{2}}c & T^{-2}d \end{bmatrix}^{-1} \begin{bmatrix} T^{-\frac{3}{2}}e \\ T^{-2}f \end{bmatrix} = \begin{bmatrix} \frac{-de+cf}{\sqrt{T}(-ad+c^2)} \\ -\frac{ce+af}{-ad+c^2} \end{bmatrix}$$

On remarque que la loi asymptotique de l'estimateur  $\widehat{\beta}_1$  est non standard, pour autant cette estimateur demeure convergent puisque on peut montrer que la distribution  $h_1$  est équivalente à une loi normale centrée sur 0. En moyenne, l'estimateur  $\widehat{\beta}_1$  est donc nul, ce qui est logique compte tenu de la définition du modèle. Mais attention cela ne signifie pas que la réalisation de  $\widehat{\beta}_1$  est nulle. Or, l'économètre étourdi à partir de la réalisation de  $\widehat{\beta}_1$  va effectuer un test de significativité pour tester si le paramètre  $\beta_1$  est nul. Reste donc à savoir si le test de Student  $\beta_1 = 0$  ne va pas conduire l'économètre à un mauvais diagnostic si ce dernier continue d'appliquer la théorie standard et par exemple le seuil de 1.96 à 5%. Par la suite, on note  $h_0$  la distribution asymptotique de  $T^{-\frac{1}{2}}\widehat{\beta}_0$ .

Déterminons donc à présent la distribution de la statistique de Student, noté  $t_{\widehat{\beta}_1}$ , associée au test de l'hypothèse  $\beta_1 = 0$ , sachant que sous (2.3.1) cette hypothèse est vraie. De façon standard, la statistique  $t_{\widehat{\beta}_1}$  est définie de la façon suivante :

$$t_{\widehat{\beta}_1} = \frac{\widehat{\beta}_1}{s_T (\xi_{22})^{\frac{1}{2}}} \quad (2.70)$$

où  $s_T^2 = \left( \sum_{t=1}^T \widehat{\eta}_t^2 \right) / (T - 2)$  désigne l'estimateur de la variance des résidus  $\eta_t$  et où  $\xi_{22}$  désigne l'élément de la deuxième colonne, deuxième ligne de la matrice  $(XX)^{-1}$ . En utilisant l'écriture (2.60), on montre que :

$$t_{\widehat{\beta}_1} = \left( \frac{\widehat{\beta}_1}{s_T T^{\frac{1}{2}}} \right) \left[ T \sum_{t=1}^T x_t^2 - \left( \sum_{t=1}^T x_t \right)^2 \right]^{\frac{1}{2}} \quad (2.71)$$

Commençons par déterminer la distribution asymptotique de  $s_T^2$  et en particulier de la somme des carrés des résidus  $SCR_T = \left( \sum_{t=1}^T \widehat{\eta}_t^2 \right)$ .

$$\begin{aligned} SCR_T &= \sum_{t=1}^T \widehat{\eta}_t^2 = \sum_{t=1}^T \left( y_t - \widehat{\beta}_0 - \widehat{\beta}_1 x_t \right)^2 \\ &= \sum_{t=1}^T y_t^2 - \sum_{t=1}^T \left( \widehat{\beta}_0 + \widehat{\beta}_1 x_t \right)^2 - 2 \sum_{t=1}^T y_t \left( \widehat{\beta}_0 + \widehat{\beta}_1 x_t \right) \end{aligned} \quad (2.72)$$

On sait que le premier élément de cette somme  $\sum_{t=1}^T y_t^2$  est  $Op(T^2)$ . Dans le second élément, la quantité  $\sum_{t=1}^T \widehat{\beta}_1^2 x_t^2$  est aussi  $Op(T^2)$ , ainsi que la quantité  $\sum_{t=1}^T \widehat{\beta}_0 x_t$ . Enfin, dans le troisième terme, seul la quantité  $\sum_{t=1}^T x_t y_t$  est  $Op(T^2)$ . Il faut donc diviser la somme des carrés des résidus par  $T^2$  pour obtenir une distribution non divergente. En regroupant les termes croisés ayant une vitesse de convergence égale à  $T^2$ , on obtient :

$$T^{-2} SCR_T = T^{-2} \left[ \sum_{t=1}^T y_t^2 - \widehat{\beta}_1^2 \sum_{t=1}^T x_t^2 - 2\widehat{\beta}_0 \sum_{t=1}^T x_t - 2\widehat{\beta}_1 \sum_{t=1}^T x_t y_t \right] + T^{-2+\alpha} Op(\alpha)$$



avec  $\alpha < 2$ . En utilisant les résultats précédents, on montre que :

$$T^{-1}s_T^2 = \frac{SCR_T}{T(T-2)} \xrightarrow[T \rightarrow \infty]{\mathcal{L}} \Upsilon \quad (2.73)$$

avec

$$\begin{aligned} \Upsilon = & \sigma_\mu^2 \int_0^1 [W_2(r)]^2 dr - h_1 \sigma_\varepsilon^2 \int_0^1 [W_1(r)]^2 dr \\ & - 2h_0 \sigma_\varepsilon \int_0^1 W_1(r) dr - 2h_1 \sigma_\varepsilon \sigma_\mu \int_0^1 W_1(r) W_2(r) dr \end{aligned} \quad (2.74)$$

où  $h_0$  et  $h_1$  désignent les distributions asymptotiques de  $\widehat{\beta}_0$  et de  $\widehat{\beta}_1$ , et où  $W_1(\cdot)$  et  $W_2(\cdot)$  sont deux mouvement Browniens standards indépendants. D'ores et déjà, on peut faire la remarque suivante :

**Remarque** Sous les hypothèses (2.3.1), l'estimateur  $s_T^2$  de la variance des résidus  $\eta_t$  dans le modèle

$$y_t = \beta_0 + \beta_1 x_t + \eta_t \quad (2.75)$$

diverge, puisque :

$$T^{-1}s_T^2 \xrightarrow[T \rightarrow \infty]{\mathcal{L}} \Upsilon \quad (2.76)$$

avec

$$\begin{aligned} \Upsilon = & \sigma_\mu^2 \int_0^1 [W_2(r)]^2 dr - h_1 \sigma_\varepsilon^2 \int_0^1 [W_1(r)]^2 dr \\ & - 2h_0 \sigma_\varepsilon \int_0^1 W_1(r) dr - 2h_1 \sigma_\varepsilon \sigma_\mu \int_0^1 W_1(r) W_2(r) dr \end{aligned} \quad (2.77)$$

où  $h_0$  et  $h_1$  désignent les distributions asymptotiques de  $\widehat{\beta}_0$  et de  $\widehat{\beta}_1$ , et où  $W_1(\cdot)$  et  $W_2(\cdot)$  sont deux mouvement Browniens standards indépendants.

Cette remarque signifie que si l'économètre étourdi ne déflate pas l'estimateur  $s_T^2$  par  $T$ , il obtiendra une réalisation d'un processus divergent. Plus  $T$  sera important, plus  $s_T^2$  sera grand. Pour  $T$  tendant vers l'infini,  $s_T^2$  diverge. Le problème c'est que  $s_T^2$  intervient dans la construction de la statistique de Student. Dès lors, nous allons montrer que cette statistique diverge elle aussi.

En effet, reprenons l'expression (2.71) de la statistique de Student

$$t_{\widehat{\beta}_1} = \left( \frac{\widehat{\beta}_1}{s_T T^{\frac{1}{2}}} \right) \left[ T \sum_{t=1}^T x_t^2 - \left( \sum_{t=1}^T x_t \right)^2 \right]^{\frac{1}{2}} \quad (2.78)$$

On sait que  $s_T$  est  $Op(T^{1/2})$ ,  $\widehat{\beta}_1$  est  $Op(1)$ ,  $\sum_{t=1}^T x_t^2$  est  $Op(T^2)$  et que  $\sum_{t=1}^T x_t$  est  $Op(T^{3/2})$ . En utilisant ces propriétés, on contrôle la vitesse de convergence des différents éléments de la façon

suivante :

$$\begin{aligned}
 t_{\hat{\beta}_1} &= \left[ \frac{\hat{\beta}_1}{(T^{-1/2} s_T) T^{1/2} T^{1/2}} \right] \left[ T T^2 \left( T^{-2} \sum_{t=1}^T x_t^2 \right) - T^3 \left( T^{-3/2} \sum_{t=1}^T x_t \right)^2 \right]^{\frac{1}{2}} \\
 &= \frac{T^{3/2}}{T} \left[ \frac{\hat{\beta}_1}{(T^{-1/2} s_T)} \right] \left[ \left( T^{-2} \sum_{t=1}^T x_t^2 \right) - \left( T^{-3/2} \sum_{t=1}^T x_t \right)^2 \right]^{\frac{1}{2}} \quad (2.79)
 \end{aligned}$$

Dès lors, on constate immédiatement que pour obtenir une valeur non divergente de la statistique de Student, celle-ci doit impérativement être multiplié par un facteur  $T^{-1/2}$ . Si l'on ne multiplie pas la statistique de Student par ce facteur, celle-ci diverge. Plus l'échantillon est grand plus les réalisations de la statistique de Student seront importantes. En effet, en utilisant les résultats asymptotiques précédents, on établit immédiatement la proposition suivante.

**Proposition 15** *Sous les hypothèses (2.3.1), la statistique de Student associée au test  $\beta_1 = 0$ , fondée sur l'estimateur des MCO, dans le modèle :*

$$y_t = \beta_0 + \beta_1 x_t + \eta_t \quad (2.80)$$

*est divergente. On montre que :*

$$\begin{aligned}
 T^{-\frac{1}{2}} t_{\hat{\beta}_1} &= \left[ \frac{\hat{\beta}_1}{(T^{-1/2} s_T)} \right] \left[ \left( T^{-2} \sum_{t=1}^T x_t^2 \right) - \left( T^{-3/2} \sum_{t=1}^T x_t \right)^2 \right]^{\frac{1}{2}} \\
 &\xrightarrow[T \rightarrow \infty]{\mathcal{L}} \left( \frac{h_1 \sigma_\varepsilon}{\Upsilon} \right) \left[ \int_0^1 [W_1(r)]^2 dr - \left( \int_0^1 W_1(r) dr \right)^2 \right] \quad (2.81)
 \end{aligned}$$

où  $h_0$  et  $h_1$  désignent les distributions asymptotiques de  $\hat{\beta}_0$  et de  $\hat{\beta}_1$ , et où  $W_1(\cdot)$  et  $W_2(\cdot)$  sont deux mouvements Browniens standards indépendants.

Ainsi, si notre économètre étourdi ne contrôle pas la vitesse de convergence de la statistique de Student, c'est à dire s'il considère uniquement les réalisations de  $t_{\hat{\beta}_1}$ , il obtiendra dès lors des réalisations d'un processus qui diverge à la vitesse  $T^{1/2}$ . Pour un même tirage de chocs, plus la taille d'échantillon sera importante, plus la réalisation de  $t_{\hat{\beta}_1}$  sera grande. Dès lors, si l'économètre applique le seuil de significativité standard de 1.96 à 5%, il aura de forte chance de rejeter à tort la nullité du paramètre  $\beta$ , et ce d'autant plus que la taille d'échantillon est importante. Ainsi si l'on régresse deux marches aléatoires qui n'ont aucun lien, les tests usuels conduisent dans de très nombreux cas à accepter à tort la significativité de la variable explicative. C'est que l'on a pu observer dans le tableau (1.6) où l'on constate que la statistique de Student est égale à  $-20$  ( $T = 1000$ ) alors que les deux séries n'ont aucun lien. Voilà pourquoi, si l'on s'en tient à la théorie usuelle des tests, la production de sous vêtements féminins au Burkina Faso est une variable explicative très importante dans la détermination du cours de l'action Microsoft sur la place de New York...

### 2.3.2 Exercices de Simulations

Afin de mieux comprendre la divergence de la statistique dans le cas d'une régression fallacieuse, nous allons à présent mener un exercice de simulation. On simule 1000 réalisations des processus  $\{x_t\}_{t=1}^T$  et  $\{y_t\}_{t=1}^T$ , pour une taille d'échantillon  $T$ , selon le modèle :

$$x_t = x_{t-1} + \varepsilon_t \quad (2.82)$$

$$y_t = y_{t-1} + \mu_t \quad (2.83)$$

avec  $x_0 = y_0 = 0$ , et  $\varepsilon_t$  *i.i.d.*  $N(0, 1)$ ,  $\mu_t$  *i.i.d.*  $N(0, 1)$ . Pour chaque pseudo échantillon, on effectue la régression :

$$y_t = \beta_0 + \beta_1 x_t + \eta_t$$

et l'on stocke la réalisation de la statistique de Student  $t_{\hat{\beta}_1}$  associée au test  $\beta_1 = 0$ . On dispose ainsi de 1000 réalisations de cette statistique à partir desquelles on peut calculer un certain nombre d'indicateurs. Dans le programme, nous faisons varier la taille des échantillons qui est d'abord fixée à  $T = 5000$ , puis à  $T = 1000$ , puis  $T = 500$  et enfin  $T = 100$ . Le programme utilisé est reporté en Annexe (A.1). Les résultats sont reportés dans le tableau (2.1)

Tableau 2.1: Résultats des Simulations :  $t_{\hat{\beta}_1}$

	$T = 100$	$T = 500$	$T = 1000$	$T = 5000$
Moyenne des $t_{\hat{\beta}_1}$	-0.03	-0.45	-0.47	0.54
Médiane des $t_{\hat{\beta}_1}$	-0.17	-0.73	-0.32	4.14
Ecart Type	10.24	22.01	31.77	69.83
Moyenne des $ t_{\hat{\beta}_1} $	8.06	17.40	24.46	55.74
Skewness	0.03	-0.03	0.03	-0.13
Kurtosis	3.03	3.05	3.46	2.81
$P\left( t_{\hat{\beta}_1}  > 1.96\right)$	0.83	0.93	0.94	0.98

Les résultats sont obtenus à partir de 1000 simulations.

On observe que plus la taille d'échantillon croît, plus l'écart type des réalisations augmente tandis que la moyenne est à peu près stable. Cela signifie que plus la taille augmente, plus les réalisations sont importantes en valeur absolue, comme le confirme la quatrième ligne du tableau (2.1). La valeur absolue de  $t_{\hat{\beta}_1}$  croît avec  $T$ , ce qui confirme notre résultat théorique de divergence. C'est pourquoi, notre économètre étourdi s'il applique les théories usuelles et s'il compare la réalisation de  $|t_{\hat{\beta}_1}|$  avec le seuil de 1.96 à 5%, rejettera d'autant plus souvent l'hypothèse nulle  $\beta_1 = 0$  que la taille  $T$  sera importante. On constate, sur la dernière ligne du tableau, que le nombre de réalisations de  $|t_{\hat{\beta}_1}|$  supérieures au seuil de 1.96 croît avec  $T$ .

Les implications statistiques de la non stationnarité sont donc particulièrement fortes, puisque dans notre exercice pour  $T = 100$ , l'économètre étourdi rejettera dans près de 83% de cas l'hypothèse de nullité de  $\beta_1$ . *Ainsi, notre économètre aura 8 chances sur dix d'affirmer que la variable  $x_t$  est significative dans la régression de  $y_t$  sur  $x_t$ , alors que les deux processus n'ont aucun lien.* Encore une fois, si l'on s'en tient à la théorie usuelle des tests, la production de sous vêtements féminins au Burkina Faso est ainsi une variable explicative très importante dans la détermination du cours de l'action Microsoft sur la place de New York... Bien entendu, une telle conclusion ne serait sans doute pas acceptée, si l'on avait au préalable stationnarisé les deux séries  $x_t$  et  $y_t$  en les différenciant. *C'est pourquoi la stationnarisation des séries est une étape fondamentale de la modélisation économétrique.*

## 2.4 Conséquences d'une mauvaise stationnarisation du processus

Nous avons dans la section précédente montrer à quel point la stationnarisation des séries était importante pour éviter de se retrouver dans la situation des régressions fallacieuses. Reste à déterminer quelle est la méthode de stationnarisation appropriée suivant que la série est  $DS$  ou  $TS$ .

**Proposition 16** *Pour stationnariser un processus  $TS$ , il convient de retirer la composante déterministe  $f_t$  en régressant la série  $x_t$  sur la plan défini par les puissances de  $t$ . Pour stationnariser un processus  $DS$  d'ordre  $d$ , il convient d'appliquer le filtre  $(1 - L)^d$ .*

Par exemple, pour stationnariser le processus  $x_t = a_0 + a_1t + \varepsilon_t$ , il suffit de régresser  $x_t$  sur une constante et sur  $t$  pour obtenir un processus  $x_t - \hat{a}_0 - \hat{a}_1t$  stationnaire. Nous allons à présent étudier les conséquences d'une mauvaise stationnarisation des processus.

### 2.4.1 Conséquence sur un processus $TS$

Supposons que  $(x_t, t \in \mathbb{Z})$  soit un processus  $TS$  défini par

$$x_t = a_0 + a_1t + \varepsilon_t$$

où  $\varepsilon_t$  est un bruit blanc gaussien  $\mathcal{N}(0, \sigma_\varepsilon^2)$ . Supposons que l'on applique, à tort, au processus  $x_t$  un filtre aux différences premières. On définit le processus  $\Delta x_t$  tel que

$$\Delta x_t = (1 - L)x_t = x_t - x_{t-1}$$

On montre que

$$\Delta x_t = a_1 + \varepsilon_t - \varepsilon_{t-1}$$

Cette écriture implique que l'application d'un filtre aux différences premières au processus  $TS$   $x_t$  a conduit à introduire une racine unitaire dans la partie moyenne mobile du processus

$\Delta x_t = (1 - L)x_t$ . Vérifions si l'application d'un filtre aux différences premières nous a permis de définir un processus  $\Delta x_t$  stationnaire. Reprenons les trois conditions de la définition de la stationnarité du second ordre.

*Convergence de  $E[(\Delta x_t)^2]$*  : On montre que  $E[(\Delta x_t)^2]$  converge puisque

$$\begin{aligned} E[(\Delta x_t)^2] &= E[(x_t - x_{t-1})^2] = E[a_1 + (\varepsilon_t - \varepsilon_{t-1})]^2 \\ &= a_1^2 + E[(\varepsilon_t - \varepsilon_{t-1})^2] = a_1^2 + 2\sigma_\varepsilon^2 < \infty \end{aligned}$$

*Moment d'ordre un* : Montrons que  $E(\Delta x_t)$  est indépendant de  $t$

$$E(\Delta x_t) = a_1 \text{ indépendant de } t$$

*Moments d'ordre deux* : Déterminons la fonction génératrice d'autocovariance  $\gamma(h)$  du processus  $\Delta x_t$ .

$$\begin{aligned} \gamma(h) &= E\{[\Delta x_t - E(\Delta x_t)][\Delta x_{t-h} - E(\Delta x_{t-h})]\} \\ &= E[(\varepsilon_t - \varepsilon_{t-1})(\varepsilon_{t-h} - \varepsilon_{t-h-1})] \end{aligned}$$

D'où l'on tire que :

$$\gamma(h) = \begin{cases} 2\sigma_\varepsilon^2 & h = 0 \\ -\sigma_\varepsilon^2 & h = \{-1; 1\} \\ 0 & \text{sinon} \end{cases} \quad (2.84)$$

On vérifie que  $\gamma(h)$  est indépendant de  $t$ . Donc les trois conditions de la stationnarité du second ordre sont vérifiées. Le processus  $\Delta x_t$  est stationnaire. Attention, résultat n'est pas général puisque par exemple l'application d'un filtre à l'ordre deux  $(1 - L)^2$  aurait conduit à un processus  $\Delta^2 x_t$  non stationnaire. Mais même si le processus  $\Delta x_t$  est stationnaire, il ne correspond à un bruit blanc.

**Remarque 1. La différenciation d'un processus *TS* conduit à une autocorrélation fallacieuse du résidu du filtre.**

En effet, la fonction génératrice d'autocovariance  $\gamma(h)$  ne correspond pas à celle d'un bruit blanc. Donc le fait d'avoir différencié  $x_t$  a introduit une autocorrélation d'ordre un de l'innovation du processus  $\Delta x_t$  qui n'existait pas dans la composante stationnaire du processus  $x_t - a_0 - a_1 t$ .

#### 2.4.2 Conséquence sur un processus DS

Supposons que  $(x_t, t \in \mathbb{Z})$  soit un processus *DS* défini d'ordre un sans dérive (*Pure Random Walk*)

$$x_t = x_{t-1} + \varepsilon_t \quad (2.85)$$

où  $\varepsilon_t$  est un bruit blanc gaussien  $\mathcal{N}(0, \sigma_\varepsilon^2)$ . Supposons que l'on applique, à tort, au processus  $x_t$  une méthode de stationnarisation consistant à régresser la série  $x_t$  sur une constante et un trend déterministe  $t$ . On considère donc le modèle empirique suivant

$$x_t = \beta_0 + \beta_1 t + \mu_t \quad (2.86)$$

On devrait alors montrer que les estimateurs des MCO  $\hat{\beta}_0$  et  $\hat{\beta}_1$  convergent vers 0 et que les  $t$ -statistiques (test de Student) associées sont généralement non significatives aux seuils standards. Or nous allons voir précisément que tel n'est pas le cas. La distribution des  $t$ -statistiques est en particulier divergente et une fois contrôlée par la vitesse de convergence, cette distribution est de plus non standard.

**Remarque 2. L'extraction d'une tendance linéaire d'un processus  $DS$  conduit à créer artificiellement une forte autocorrélation des résidus aux premiers retards et donc un mouvement pseudo-périodique des résidus.**

Considérons un petit exemple. Soit le processus  $x_t = x_{t-1} + \varepsilon_t$  où  $\varepsilon_t$  suit une  $\mathcal{N}(0, 1/4)$ . On considère 1000 échantillons du bruit blanc  $\varepsilon_t$  à partir desquels sont construites 1000 pseudo séries  $x_t$ . Pour chacun de ces pseudo échantillons, on étudie la régression :

$$x_t = \beta_0 + \beta_1 t + \mu_t$$

On obtient alors les résultats suivants :

- moyenne des  $\hat{\beta}_0 = 1.3081$  moyenne des  $\hat{\beta}_1 = -0.0007$ .  $var(T^{\frac{1}{2}}\hat{\beta}_1) \simeq \frac{4}{3}\sigma_\varepsilon^2$
- moyenne des  $\hat{t}_{\beta_0} = 5.26$  et moyenne des  $\hat{t}_{\beta_1} = -2.34$
- $P(|\hat{t}_{\beta_0}| > 1.96) = 93\%$  et  $P(|\hat{t}_{\beta_1}| > 1.96) = 96\%$ . Cela signifie que si l'on applique à tort les seuils de significativité standard, dans plus de 90% des cas on rejette l'hypothèse de nullité des coefficients  $\beta_0$  et  $\beta_1$ .
- Autocorrélation moyenne des résidus  $\hat{\mu}_t$  : ordre 1 : 0.9899, ordre 2 : 0.9799, ordre 3 : 0.9699, ordre 10 : 0.9012.

Tout ceci montre l'importance de bien choisir la méthode de stationnarisation des séries en fonction de l'origine de la non stationnarité :  $DS$  ou  $TS$  Il convient donc à présent de présenter des tests qui nous permettent, tout d'abord de vérifier que les séries sont non stationnaires et d'autre part de discriminer entre les processus  $DS$  et  $TS$  : ce sont les tests de racine unitaire.

### 3 Un Test de Racine Unitaire : le Test de Dickey Fuller

Le test de Dickey Fuller simple (1979) est un *test de racine unitaire (ou de non stationnarité) dont l'hypothèse nulle est la non stationnarité d'un processus autorégressif d'ordre un*. Considérons un processus  $(x_t, t \in \mathbb{Z})$  satisfaisant la représentation  $AR(1)$  suivante :

$$x_t = \rho x_{t-1} + \varepsilon_t \quad (3.87)$$

avec  $\varepsilon_t$  *i.i.d.*  $(0, \sigma_\varepsilon^2)$ , et  $\rho \in \mathbb{R}$ . Le principe général du test de Dickey Fuller consiste à tester l'hypothèse nulle de la présence d'une racine unitaire :

$$H_0 : \rho = 1 \quad (3.88)$$

$$H_a : |\rho| < 1 \quad (3.89)$$

En effet, sous l'hypothèse nulle  $H_0$ , le processus (3.87) se ramène à une pure marche aléatoire (*Random Walk Process*). *L'hypothèse nulle testée correspond ainsi à une hypothèse de non stationnarité stochastique*<sup>6</sup>. Ce test, comme tout test non symétrique, peut être réalisé de différentes façons. La plus simple consiste à utiliser une statistique de Student associée à l'hypothèse  $H_0$ . En cela, le test de Dickey Fuller ne se distingue pas d'un test quelconque d'une hypothèse non symétrique. Là où l'application du test de Dickey Fuller diffère de celle d'un test standard, c'est dans la distribution asymptotique de la statistique de Student associée au test  $H_0$ . En effet, *nous allons montrer que la distribution asymptotique de l'estimateur des MCO du paramètre  $\rho$ , sous l'hypothèse de non stationnarité, est non standard*. On n'a plus dans ce cas une distribution asymptotique normale, comme en économétrie de base. De la même façon, la statistique de Student associée au tests  $\rho = 1$ , n'a pas une distribution asymptotique standard (distribution de Student approximée par une distribution normale). *C'est pourquoi, l'application du test de Dickey Fuller nécessite que l'on utilise des seuils différents de ceux que l'on utilise traditionnellement pour des statistiques de Student*.

De plus, puisqu'un malheur n'arrive jamais seul, *nous montrerons que la distribution asymptotique de la statistique de Student associée au test  $H_0$ , n'est pas la même suivant que dans le modèle (3.87) on inclut ou non, une constante et un trend déterministe*. Dès lors, puisque a priori, on se sait pas si l'on doit inclure cette constante et ce trend, il convient d'appliquer non pas un test simple, mais une **stratégie de tests de Dickey Fuller**.

---

<sup>6</sup>C'est pourquoi le test de Dickey Fuller, contrairement à d'autres tests que nous verrons par la suite, est un *test de non stationnarité*, et non pas un test de stationnarité.

### 3.1 Des distributions asymptotiques non standard sous $H_0$

Considérons le modèle le plus simple sans constante, ni trend. Soit un processus  $(x_t, t \in \mathbb{Z})$  satisfaisant la représentation  $AR(1)$  suivante :

$$x_t = \rho x_{t-1} + \varepsilon_t \quad (3.90)$$

avec  $\varepsilon_t$  *i.i.d.*  $(0, \sigma_\varepsilon^2)$ . Dans ce modèle, le test de Dickey Fuller revient à tester l'hypothèse nulle  $\rho = 1$ . Nous commencerons par étudier la distribution asymptotique de l'estimateur des MCO du paramètre  $\rho$ , sous l'hypothèse  $H_0$ . Puis nous étudierons la distribution la statistique de Student associée au test  $\rho = 1$ .

#### 3.1.1 Distribution de l'estimateur $\hat{\rho}$

D'après la définition des MCO, pour un échantillon de taille  $T$ , l'estimateur  $\hat{\rho}$  est défini par :

$$\hat{\rho} = \frac{\sum_{t=1}^T x_t x_{t-1}}{\sum_{t=1}^T x_{t-1}^2} \quad (3.91)$$

Nous savons que sous, l'hypothèse  $H_0$  ( $\rho = 1$ ), le processus générateur de données (modèle 3.108) se ramène alors à une simple marche aléatoire

$$x_t = x_{t-1} + \varepsilon_t \quad (3.92)$$

avec  $\varepsilon_t$  *i.i.d.*  $(0, \sigma_\varepsilon^2)$ . Ainsi on a  $\sum_{t=1}^T x_t x_{t-1} = \sum_{t=1}^T x_{t-1}^2 + \sum_{t=1}^T x_{t-1} \varepsilon_t$ . Dès lors, le biais associé à l'estimateur  $\hat{\rho}$  peut s'écrire sous la forme :

$$\hat{\rho} - 1 = \frac{\sum_{t=1}^T x_{t-1} \varepsilon_t}{\sum_{t=1}^T x_{t-1}^2} \quad (3.93)$$

Afin de contrôler les vitesses de convergence des moments empiriques apparaissant au numérateur et au dénominateur, nous allons transformer cette expression de la façon suivante :

$$T(\hat{\rho} - 1) = \frac{T^{-1} \sum_{t=1}^T x_{t-1} \varepsilon_t}{T^{-2} \sum_{t=1}^T x_{t-1}^2} \quad (3.94)$$

Nous savons que sous, l'hypothèse  $H_0$  ( $\rho = 1$ ), le processus générateur de données est une simple marche aléatoire. En appliquant le *Théorème Central Limite Fonctionnel* et le *Continuous Mapping Theorem*, on sait que dans ce cas :

$$T^{-1} \sum_{t=1}^T x_{t-1} \varepsilon_t \xrightarrow[T \rightarrow \infty]{\mathcal{L}} \frac{\sigma_\varepsilon^2}{2} [W(1)^2 - 1] \quad (3.95)$$



$$T^{-2} \sum_{t=1}^T x_{t-1}^2 \xrightarrow[T \rightarrow \infty]{\mathcal{L}} \sigma_\varepsilon^2 \int_0^1 [W(r)]^2 dr \quad (3.96)$$

où  $W(\cdot)$  désigne un mouvement Brownien scalaire standard. On peut alors établir la distribution asymptotique de l'estimateur  $\hat{\rho}$  sous l'hypothèse  $H_0$ .

**Proposition 17** *Sous l'hypothèse  $H_0$  de non stationnarité,  $\rho = 1$ , la distribution asymptotique de l'estimateur  $\hat{\rho}$  des MCO du paramètre  $\rho$  dans le modèle*

$$x_t = \rho x_{t-1} + \varepsilon_t \quad (3.97)$$

avec  $\varepsilon_t$  i.i.d.  $(0, \sigma_\varepsilon^2)$ , est la suivante :

$$T(\hat{\rho} - 1) \xrightarrow[T \rightarrow \infty]{\mathcal{L}} \frac{1}{2} \frac{[W(1)]^2 - 1}{\int_0^1 [W(r)]^2 dr} \quad (3.98)$$

où  $W(\cdot)$  est un mouvement Brownien standard.

On vérifie ainsi que conformément à ce que nous avons énoncé en introduction, la distribution asymptotique de l'estimateur  $\hat{\rho}$  obtenue sous  $H_0$  est non standard, et en particulier non symétrique. Plusieurs remarques peuvent être faites à ce niveau, qui nous permettront de bien comprendre les implications de la non stationnarité.

**Remarque 1.** L'estimateur  $\hat{\rho}$  converge à la vitesse  $1/T$ , sous l'hypothèse  $H_0$   $\rho = 1$ , alors qu'il converge à la vitesse  $1/T^{1/2}$  sous l'hypothèse alternative  $H_a$ ,  $|\rho| < 1$  :

$$\text{Sous } H_0 : T(\hat{\rho} - 1) \xrightarrow[T \rightarrow \infty]{\mathcal{L}} \frac{1}{2} \frac{[W(1)]^2 - 1}{\int_0^1 [W(r)]^2 dr} \quad (3.99)$$

$$\text{Sous } H_a : \sqrt{T}(\hat{\rho} - \rho) \xrightarrow[T \rightarrow \infty]{\mathcal{L}} N\left(0, \sigma_\phi^2\right) \quad (3.100)$$

Une autre façon de comprendre ce résultat est la suivante. *Quelle que soit l'hypothèse retenue, stationnarité ou non, l'estimateur des MCO converge en probabilité vers la vraie valeur  $\rho$  (avec  $\rho = 1$ , si  $H_0$  est vraie,  $|\rho| < 1$ , si  $H_a$  est vraie).*

$$\hat{\rho} \xrightarrow[T \rightarrow \infty]{P} \rho$$

Toutefois, si les données sont non stationnaires ( $H_0 : \rho = 1$ ), l'estimateur des MCO, converge "plus vite" vers la vraie valeur du paramètre  $\rho$ , égale à 1 dans ce cas là. La variance de la distribution asymptotique "s'écrase" alors à la vitesse  $T$  sous  $H_0$ , alors qu'elle ne s'écrase qu'à la vitesse  $\sqrt{T}$  dans le cas stationnaire (cf. illustration de la convergence en probabilité, chapitre 1).

**Remarque 2.** Sous l'hypothèse de non stationnarité  $H_0$ , la distribution asymptotique de l'estimateur  $\hat{\rho}$  est non standard (non normale). En particulier, cette distribution est non symétrique.

Rappelons que dans le cas stationnaire, la distribution asymptotique de  $\sqrt{T}(\hat{\rho} - \rho)$  est une loi normale centrée, donc par définition une distribution symétrique. En revanche, sous  $H_0$ , la distribution de  $T(\hat{\rho} - 1)$  n'est pas une distribution normale, et en particulier cette distribution est non symétrique.

$$T(\hat{\rho} - 1) \xrightarrow[T \rightarrow \infty]{\mathcal{L}} \frac{1}{2} \frac{[W(1)]^2 - 1}{\int_0^1 [W(r)]^2 dr}$$

En effet, étant donnée la définition d'un mouvement Brownien standard, le processus  $W(1)^2$  suit approximativement un  $\chi^2(1)$ . Sachant que  $P[\chi^2(1) < 1] = 0.68$ , cela signifie que la distribution de  $T(\hat{\rho} - 1)$ , conditionnellement à  $\int_0^1 [W(r)]^2 dr$ , est non symétrique et qu'environ 2/3 des réalisations de  $\hat{\rho}$  seront inférieures à 1.

Pour illustrer cette dernière remarque, on peut réaliser un *exercice de simulation* sous Eviews. L'idée générale de cet exercice est la suivante. On va répliquer, par la méthode de Monte Carlo,  $N$  échantillons (qualifiés de *pseudo échantillons*) du processus  $x_t$  défini par l'équation (3.90) sous l'hypothèse nulle  $\rho = 1$ . Pour cela il suffit de se donner une valeur numérique pour  $\sigma_\varepsilon^2$  et de tirer  $N$  échantillons de taille  $T$  dans une loi normale<sup>7</sup>  $N(0, \sigma_\varepsilon^2)$ . A partir de ces  $N$  échantillons du processus  $\{\varepsilon_t\}_{t=1}^T$ , on reconstruit  $N$  échantillons de taille  $T$  du processus  $\{x_t\}_{t=1}^T$  selon l'équation  $x_t = x_{t-1} + \varepsilon_t$ , en se donnant pour chaque échantillon une condition initiale sur  $x_0$  (par exemple  $x_0 = 0$ ). Ensuite, il suffit d'effectuer la régression de  $x_t$  sur  $x_{t-1}$  et de stocker pour chaque échantillon la réalisation de l'estimateur  $\hat{\rho}$  des MCO. Ainsi, au final, on obtient  $N$  réalisations de cet estimateurs, à partir desquelles on peut calculer un certain nombre de statistiques et construire, en particulier, un estimateur de la distribution de  $\hat{\rho}$ .

Tableau 3.2: Valeurs des Paramètres de la Simulation

$N$	$T$	$\sigma_\varepsilon^2$	$x_0$
100	5000	1	0

Pour approcher au mieux de la distribution asymptotique il convient de choisir une taille  $T$  d'échantillon assez importante. De plus, il convient de faire un grand nombre  $N$  de simulations pour obtenir des statistiques indépendantes des tirages effectués dans la loi normale. Les paramètres que nous avons retenus pour cet exercice sont reportés dans le tableau (3.2).

<sup>7</sup>L'hypothèse de normalité des  $\varepsilon_t$  est en effet compatible avec l'hypothèse  $\varepsilon_t$  *i.i.d.* de notre modèle.

Le programme<sup>8</sup> utilisé sous Eviews est le suivant :

```

'- VECTEUR DES REALISATIONS DE L ESTIMATEUR DES MCO-
VECTOR (100) RHO
FOR !I = 1 TO 100
    '-INITIALISATION DU PROCESSUS-
    SMPL 1 5001
    GENR X = 0
    '-CONSTRUCTION DE LA MARCHE ALEATOIRE-
    GENR EPS = NRND
    SMPL 2 5001
    GENR X = X(-1)+EPS
    '-REGRESSION OLS-
    EQUATION EQ1.LS X X(-1)
    RHO(!I) = EQ1.@COEFS(1)
NEXT
'-CONVERSION DES VECTEURS EN SERIES
SMPL 1 100
MTOS(RHO,RHO-S)
GENR BIAIS-S = (RHO-S-1)*5000

```

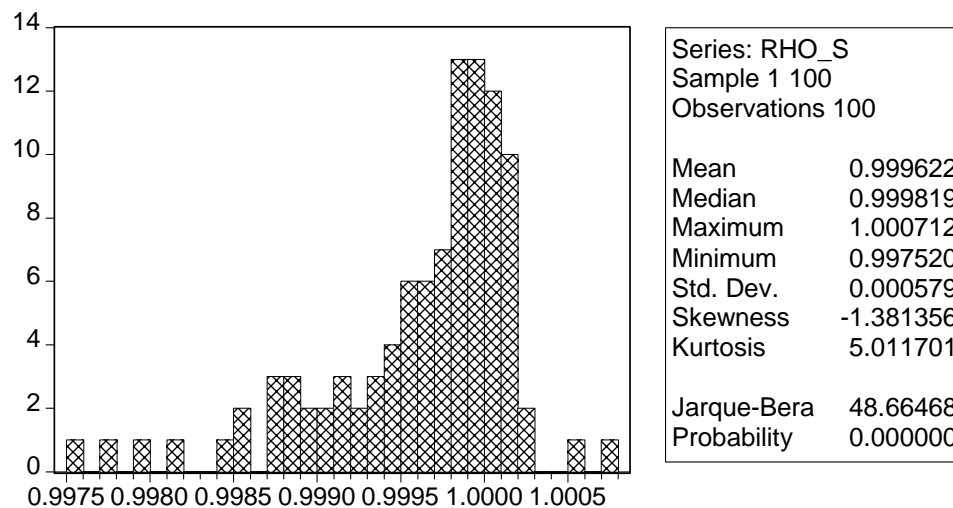
Ce programme débute par la déclaration d'un vecteur, nommé RHO, de dimension  $(N, 1)$  dans lequel seront stockées les valeurs de la réalisation de l'estimateur  $\hat{\rho}$  pour chaque simulation. Ensuite une boucle est construite avec un indice  $i$  allant de 1 à  $N = 100$ . Au sein de cette boucle, on commence par initialiser le processus  $x_t$  en l'égalisant à 0 sur l'ensemble de la période allant de 1 à 5001 (la première valeur correspondant au  $x_0$ ). Ensuite, à l'aide de la commande NRND on effectue un tirage dans une loi normale  $N(0, 1)$  que l'on stocke dans une série EPS. En réduisant la période d'échantillon, de la date 2 à la date 5001, on reconstruit la série  $x_t$  selon le modèle  $x_t = x_{t-1} + \varepsilon_t$ . La valeur retardée de  $x_t$  se note  $x(-1)$ . Reste enfin à effectuer la régression de  $x_t$  sur  $x_{t-1}$  (LS X X(-1)) et à stocker dans le vecteur RHO la valeur de l'estimateur  $\hat{\rho}$  obtenue pour le  $i^{\text{ème}}$  pseudo échantillon, grâce à l'instruction  $\text{RHO}(!I) = \text{EQ1}.\text{@COEFS}(1)$ . Au sortir de la boucle, on convertit le vecteur RHO en une série RHO-S afin d'effectuer plus facilement certains calculs, et l'on construit la variable transformée BIAIS-S correspondant à la v.a.r.  $T(\hat{\rho} - 1)$ .

En cliquant sur la série RHO-S, puis en cliquant sur *View, Descriptive Statistics, et Histogramm and Stats*, on obtient les informations qui sont reportées sur la figure (3.8). Tout d'abord, ces

---

<sup>8</sup>Pour lancer ce programme, il convient au préalable de créer un *Workfile* de type *Undated or Irregular*, de 1 à 5001, puis de charger le programme (*open program*) ou de le taper (*new program*), et ensuite de l'exécuter (*Run*).

résultats apparaissent compatibles avec le fait que l'estimateur  $\hat{\rho}$  soit un estimateur convergent ( $\text{plim } \hat{\rho} = 1$ ) puisque l'on observe que la moyenne des 100 réalisations des  $\hat{\rho}$  est très proche de 1. La distribution des  $\hat{\rho}$  apparaît en outre relativement concentrée autour de la vraie valeur  $\rho = 1$ . Mais ce qui est le plus frappant, c'est que l'on vérifie à partir de cette petite expérience *la non symétrie de la distribution* (cf. remarque 2). On peut le vérifier visuellement à partir de l'histogramme, mais en outre cette intuition est confortée par l'examen de la Skewness qui vaut  $-1.38$  dans cette expérience et qui est donc largement différente de 0. Le test de normalité de Jarque-Bera (cf. chapitre 4) rejette ainsi la normalité de la distribution.

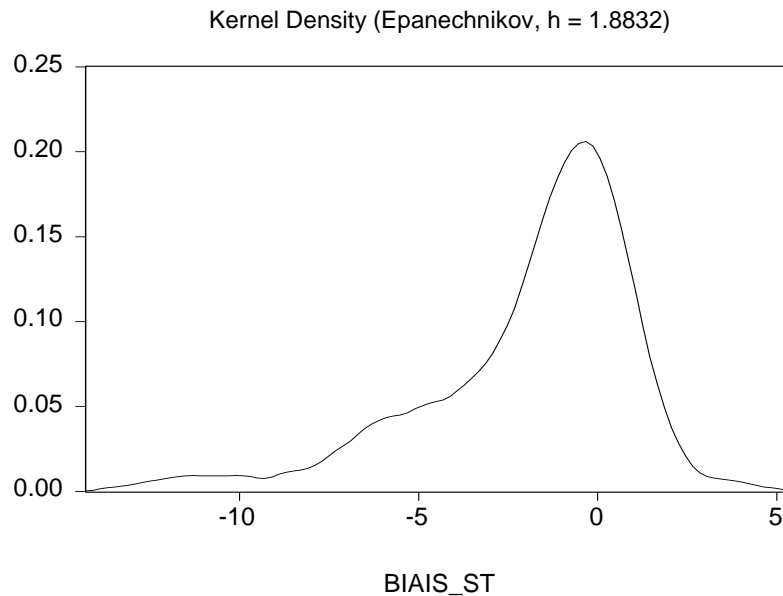
Figure 3.8: Histogramme et Statistiques Descriptives des  $\hat{\rho}$ 

Ce premier examen quand à la non normalité de la distribution, est toute fois incomplet et non fondé sur le plan théorique. En effet, nous avons montré que la distribution asymptotique de  $\hat{\rho}$  était dégénérée dans le sens où  $\hat{\rho}$  converge en probabilité vers 1. Ainsi si l'on fait tendre  $T$  vers l'infini, la distribution des  $\hat{\rho}$  convergera vers une masse ponctuelle en 1. Donc pour étudier correctement la forme de la distribution asymptotique des  $\hat{\rho}$ , il convient d'étudier le comportement de la transformée  $T(\hat{\rho} - 1)$  qui elle possède une distribution non dégénérée. Pour ce faire, on clique sur la série BIAIS-S, puis sur *View, Distribution Graphs, Kernel Density*. On obtient alors un estimateur (*estimateur à noyau*) de la fonction de densité théorique de la v.a.r.  $T(\hat{\rho} - 1)$ . Cet estimateur de la densité théorique est reproduit sur la figure (3.9).

Sur la figure (3.9), on vérifie bien que la distribution asymptotique de  $T(\hat{\rho} - 1)$  n'est pas une loi normale, conformément à la proposition (17). On vérifie en outre que cette distribution est non symétrique, et que la probabilité d'obtenir une valeur de  $\hat{\rho}$  inférieure à 1, est largement supérieure à celle d'obtenir une valeur supérieure, et cela dans un rapport de  $2/3, 1/3$ . En effet, la surface définie

par la fonction de distribution empirique de  $T(\hat{\rho} - 1)$  entre  $-\infty$  et 1 est largement supérieure à celle définie entre 1 et  $+\infty$ . Ainsi pour les  $N = 100$  simulations effectuées, nous avons obtenu 74 réalisations inférieures à 1, et 26 supérieures. La probabilité empirique 0.74 d'obtenir une valeur inférieure à 1 est donc relativement proche de la probabilité théorique de 0.68. Pour un nombre de simulations  $N$  plus important, la différence aurait été encore plus faible.

Figure 3.9: Estimateur de la Fonction de Densité Théorique des  $T(\hat{\rho} - 1)$



### 3.1.2 Distribution de la statistique de Student $t_{\hat{\rho}=1}$

Nous allons maintenant nous intéresser au coeur du problème du test de Dickey Fuller, à savoir le comportement asymptotique de la statistique de Student associée au test  $\rho = 1$  dans le modèle :

$$x_t = \rho x_{t-1} + \varepsilon_t \quad (3.101)$$

avec  $\varepsilon_t$  *i.i.d.*  $(0, \sigma_\varepsilon^2)$ . La question que l'on se pose sur le plan pratique est la suivante : *peut-on, sous l'hypothèse  $H_0$ , adopter les seuils standard d'une loi normale (1.96 à 5% par exemple) pour effectuer le test de Dickey Fuller ?* Exprimée en d'autres termes, cette question revient à déterminer si la distribution asymptotique de la statistique de Student associée au test  $\rho = 1$ , correspond, sous  $H_0$ , à une loi de Student, qui peut être approximée par une loi normale pour  $T \rightarrow \infty$  (cf. chapitre 1). Si tel n'est pas le cas, alors le test de Dickey Fuller ne devra bien entendu pas être construit à partir des seuils standards de la loi normale.

Pour répondre à cette question, commençons par définir la statistique de Student sur laquelle sera fondée le test de Dickey Fuller de l'hypothèse nulle  $\rho = 1$ . Cette dernière, notée  $t_{\hat{\rho}=1}$ , est tout simplement définie de la façon suivante :

$$t_{\hat{\rho}=1} = \frac{(\hat{\rho} - 1)}{\sigma_{\hat{\rho}}} = (\hat{\rho} - 1) \left[ \frac{\sum_{t=1}^T x_{t-1}^2}{s_T^2} \right]^{\frac{1}{2}} \quad (3.102)$$

où  $s_T^2$  désigne l'estimateur empirique de la variance des résidus :

$$s_T^2 = \frac{1}{(T-1)} \sum_{t=1}^T \hat{\varepsilon}_t^2 = \frac{1}{(T-1)} \sum_{t=1}^T (x_t - \hat{\rho}x_{t-1})^2 \quad (3.103)$$

Afin de contrôler la vitesse de convergence des moments intervenant au numérateur et au dénominateur de  $t_{\hat{\rho}=1}$ , nous allons étudier la distribution asymptotique de  $t_{\hat{\rho}=1}$  exprimée sous la forme suivante :

$$t_{\hat{\rho}=1} = \frac{T^{-1} \sum_{t=1}^T x_{t-1} \varepsilon_t}{\left[ T^{-2} \sum_{t=1}^T x_{t-1}^2 \right]^{\frac{1}{2}} s_T} \quad (3.104)$$

Essayons à présent d'établir la distribution asymptotique de  $t_{\hat{\rho}=1}$  sous  $H_0$ . Pour cela étudions tout d'abord le comportement asymptotique de  $s_T^2$ . On sait que sous  $H_0$ , ou sous  $H_a$ , l'estimateur  $\hat{\rho}$  est convergent ( $plim \hat{\rho} = \rho$ ). Donc par construction, les résidus estimés  $\hat{\varepsilon}_t = x_t - \hat{\rho}x_{t-1}$  sont non biaisés. Dans ce cas, on montre que le moment empirique d'ordre deux  $s_T^2$  converge en probabilité vers la variance  $\sigma_\varepsilon^2$  de la population des résidus  $\varepsilon_t$  (cf. chapitre 1).

$$\hat{\phi} \xrightarrow[T \rightarrow \infty]{p} \phi \implies s_T^2 \xrightarrow[T \rightarrow \infty]{p} \sigma_\varepsilon^2 \quad (3.105)$$

Reprenons maintenant un à un les différents éléments de l'expression de  $t_{\hat{\rho}=1}$  (équation 3.104) et étudions leurs distributions respectives. D'après les résultats de la première section, on a :

$$\begin{aligned} T^{-1} \sum_{t=1}^T x_{t-1} \varepsilon_t &\xrightarrow[T \rightarrow \infty]{\mathcal{L}} \frac{\sigma_\varepsilon^2}{2} [W(1)^2 - 1] \\ T^{-2} \sum_{t=1}^T x_{t-1}^2 &\xrightarrow[T \rightarrow \infty]{\mathcal{L}} \sigma_\varepsilon^2 \int_0^1 [W(r)]^2 dr \\ s_T &\xrightarrow[T \rightarrow \infty]{p} \sigma_\varepsilon \end{aligned}$$

En utilisant les différentes propriétés énumérées dans le cadre des rappels du premier chapitre sur la convergence en loi et en probabilité, on peut alors immédiatement dériver la distribution asymptotique de  $t_{\hat{\rho}=1}$ .

**Proposition 18** *Sous l'hypothèse  $H_0$  de non stationnarité, la distribution asymptotique de la statistique de Student associée au test de Dickey Fuller,  $\rho = 1$ , dans le modèle*

$$x_t = \rho x_{t-1} + \varepsilon_t \quad (3.106)$$

avec  $\varepsilon_t$  i.i.d.  $(0, \sigma_\varepsilon^2)$ , est la suivante :

$$t_{\hat{\rho}=1} \xrightarrow[T \rightarrow \infty]{\mathcal{L}} \frac{1}{2} \frac{[W(1)]^2 - 1}{\left\{ \int_0^1 [W(r)]^2 dr \right\}^{\frac{1}{2}}} \quad (3.107)$$

où  $W(\cdot)$  est un mouvement Brownien standard.

Ainsi, on vérifie que cette distribution asymptotique n'est pas une distribution normale. Encore une fois, en particulier, cette distribution est non symétrique en raison de la présence du terme  $W(1)^2$  distribué selon un  $\chi^2$ . Les seuils de cette loi ont été tabulés par Dickey et Fuller (1979) et par d'autres auteurs comme Mc Kinnon (1981).

**Remarque 1.** La distribution asymptotique, sous  $H_0$ , de la statistique de Student  $t_{\hat{\rho}=1}$  du test de Dickey Fuller dans le modèle (3.101) n'est pas standard. L'utilisation, à tort, des seuils standard associés à une distribution normale peut ainsi conduire à un mauvais diagnostic quant à la non stationnarité de la série étudiée. En particulier, ce type d'erreur conduit à rejeter trop souvent l'hypothèse de non stationnarité.

Essayons, de bien comprendre les enjeux de cette remarque. Pour cela, menons à nouveau un exercice de simulation. On simule, de la même façon que précédemment, un grand nombre ( $N$ ) de pseudo échantillons du processus  $\{x_t\}_{t=1}^T$  sous l'hypothèse  $H_0$ . Pour chaque pseudo échantillon, on réalise la régression de  $x_t$  sur  $x_{t-1}$ , et l'on conserve la réalisation de la statistique de Student  $t_{\hat{\rho}=1}$  associée au test  $\rho = 1$ . On obtient  $N$  réalisations de cette statistique, à partir desquelles on peut la distribution de  $t_{\hat{\rho}=1}$ .

Dans la pratique, les principaux logiciels d'économétrie (et en particulier Eviews), proposent une statistique de Student associée au test de la nullité d'un paramètre déjà programmée, alors que le test de l'égalité d'un paramètre à 1 n'est pas toujours programmé. C'est pourquoi, pour éviter de programmer la construction de  $t_{\hat{\rho}=1}$ , nous allons tout d'abord transformer ce modèle avant d'appliquer le test de Dickey Fuller.

**Proposition 19** *Le test de l'hypothèse  $\rho = 1$  dans l'équation (3.90) est identique au test de l'hypothèse  $\phi = 0$  dans le modèle transformé suivant :*

$$\Delta x_t = \phi x_{t-1} + \varepsilon_t \quad (3.108)$$

avec  $\phi = \rho - 1$  et  $\Delta x_t = (1 - L)x_t = x_t - x_{t-1}$ . Dans ce modèle, le test de Dickey Fuller se ramène alors à  $H_0 : \phi = \rho - 1 = 0$  contre  $H_a : \phi < 0$ . La statistique  $t_{\hat{\phi}=0}$  a la même distribution asymptotique que  $t_{\hat{\rho}=1}$ .

Pour cet exercice de simulation, nous avons retenu les mêmes paramètres que pour la simulation précédente (cf. tableau 3.2), à l'exception du nombre de simulations  $N$  que nous avons porté à 1000. Le programme<sup>9</sup> sous Eviews est le suivant :

```

!- T STATISTIQUE : TEST DE DICKEY FULLER-
VECTOR (1000) TSTAT
FOR !I = 1 TO 1000
    !-INITIALISATION DU PROCESSUS-
    SMPL 1 5001
    GENR X = 0
    !- CONSTRUCTION DE LA MARCHE ALEATOIRE-
    GENR EPS = NRND
    SMPL 2 5001
    GENR X = X(-1)+EPS
    GENR DX = X - X(-1)
    !-REGRESSION OLS-
    EQUATION EQ1.LS DX X(-1)
    TSTAT(!I) = EQ1.@TSTATS(1)
NEXT
! - Conversion des Vecteurs en Series
SMPL 1 1000
MTOS(TSTAT,TSTAT-S)

```

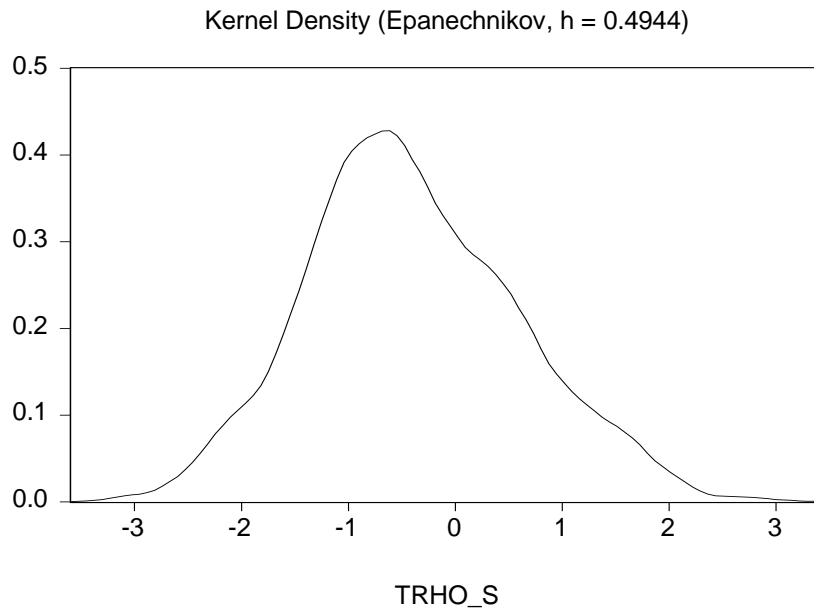
L'idée générale du programme est sensiblement la même que pour l'exercice précédent, la seule différence étant que cette fois-ci, on construit le vecteur des statistiques de Student grâce à l'instruction @TSTATS(1) qui permet de récupérer les t-stats de la régression des MCO. L'estimateur de la fonction de densité de  $t_{\hat{\phi}=0}$  obtenu à partir des 1000 simulations est reporté sur la figure (3.10). On constate une nouvelle fois que la distribution asymptotique de  $t_{\hat{\phi}=0}$  (qui a la même distribution asymptotique que  $t_{\hat{\rho}=1}$ ) n'est pas une loi normale. On vérifie, que conformément au résultat théorique, cette distribution n'est pas symétrique.

À partir des 1000 simulations, il est en outre possible de calculer les quantiles au seuil  $\alpha$  associés au test non symétrique  $\rho = 1$  et de le comparer à ceux que l'on obtiendrait, si l'on supposait, à

<sup>9</sup>Pour lancer ce programme, il convient au préalable de créer un *Workfile* de type *Undated or Irregular*, de 1 à 5001, puis de charger le programme (*open program*) ou de le taper (*new program*), et ensuite de l'exécuter (*Run*).



Figure 3.10: Estimateur de la Fonction de Densité Empirique de  $t_{\hat{\phi}=0}$



tort que  $t_{\hat{\rho}=1}$  était distribuée selon une loi  $N(0, 1)$ . On cherche donc le seuil  $C_{(\alpha)}$  tel que<sup>10</sup> :

$$P \left\{ t_{\hat{\phi}=0} \leq C_{(\alpha)} \right\} = \alpha \tag{3.109}$$

Pour cela, il suffit de classer dans l'ordre croissant les 1000 réalisations de  $t_{\hat{\phi}=0}$  et de considérer la  $N/100 \times \alpha^{\text{ème}}$  réalisation. Par exemple, pour le seuil à 5%, après avoir classé les réalisations, on retiendra la 50<sup>ème</sup> observation, puisque empiriquement, il existe 95% de chances que l'on ait une réalisation supérieure à ce seuil. Sur le tableau (3.3) sont reportés les seuils obtenus à partir des simulations, et les seuils théoriques associés à un test unilatéral fondé sur une loi  $N(0, 1)$ .

Tableau 3.3: Seuils Critiques des Tests  $t_{\hat{\phi}=0}$

	$\alpha = 1\%$	$\alpha = 5\%$	$\alpha = 10\%$
$C_{(\alpha)}$	-2,44	-1,98	-1,58
$C_{(\alpha)}^{N(0,1)}$	-2.32	-1.64	-1.28

On observe ainsi que si l'on considère un risque de première espèce de 5%, le seuil d'une loi normale pour le test unilatéral  $\rho = 1$ , est égal à  $-1.64$ . Mais, nous avons vu théoriquement que la statistique  $t_{\hat{\phi}=0}$  ne suit pas une loi normale. Les 1000 simulations montrent que le vrai seuil à

<sup>10</sup>Contrairement au test standard, on ne considère pas ici la valeur absolue de la t-stat : le test est non symétrique.

5% de la loi de  $t_{\hat{\phi}=0}$ , est en fait plus proche de  $-1.98$ . Ainsi, *on vérifie que l'on aurait eu tendance à trop souvent rejeter la non stationnarité si l'on avait appliqué pour le test de Dickey Fuller les seuils standard d'une loi normale.* Le même diagnostic vaut pour les autres niveaux de risque. C'est pourquoi, *il est nécessaire d'utiliser les seuils tabulés par Dickey Fuller (1979) ou Mc Kinnon (1981) pour effectuer les test de non stationnarité.*

### 3.2 Des distributions conditionnelles au modèle choisi...

Comme un malheur n'arrive jamais seul, le pauvre économètre qui voit déjà disparaître sa sacro sainte loi normale dans l'application des tests de Dickey Fuller, va de plus être confronté à un autre type de problème : *celui de la dépendance des distributions asymptotiques au modèle choisi.*

**Proposition 20** *Sous l'hypothèse  $H_0$  de non stationnarité, la distribution asymptotique de la statistique de Student  $t_{\hat{\rho}=1}$  diffère suivant que le modèle utilisé soit :*

$$x_t = \rho x_{t-1} + \varepsilon_t \quad (3.110)$$

$$x_t = \rho x_{t-1} + c + \varepsilon_t \quad (3.111)$$

$$x_t = \rho x_{t-1} + c + \beta t + \varepsilon_t \quad (3.112)$$

avec  $\varepsilon_t$  i.i.d.  $(0, \sigma_\varepsilon^2)$ .

Ainsi, le fait d'introduire une constante, ou une constante et un trend déterministe affecte la distribution asymptotique de  $t_{\hat{\rho}=1}$ . *Pour le praticien, cela signifie que les seuils critiques du tests de Dickey Fuller ne seront pas identiques, suivant que l'on inclut ou non dans le modèle une constante et un trend.* Cela pose naturellement un problème, puisque a priori, on ne sait pas si l'on doit inclure ces éléments dans le modèle testé. C'est pourquoi, *on propose généralement une stratégie de tests de Dickey Fuller, et non pas un seul test unique.*

Nous ne démontrerons pas le résultat général de cette proposition. Nous nous contenterons de démontrer que la distribution de l'estimateur  $\hat{\rho}$  obtenue dans le modèle (3.110) diffère de celle obtenue dans le modèle (3.111). Naturellement, il en découle que la distribution de  $t_{\hat{\rho}=1}$  dans les deux modèles diffère. Déterminons la distribution de  $t_{\hat{\rho}=1}$  à partir du modèle (3.111). Pour ce faire, nous considérerons pour simplifier que sous  $H_0$ , le processus  $x_t$  est non stationnaire (hypothèse du test de Dickey Fuller), mais qu'en plus il se ramène à une *marche aléatoire pure* ( $c = 0$ ). Ainsi supposons que les données soient générées par le modèle :

$$x_t = x_{t-1} + \varepsilon_t \quad (3.113)$$

avec  $\varepsilon_t$  *i.i.d.*  $(0, \sigma_\varepsilon^2)$ . Le modèle estimé par les MCO, pour effectuer le test de non stationnarité, inclut une constante (modèle 1 des tests Dickey Fuller donné par l'équation 3.111) :

$$x_t = \rho x_{t-1} + c + \varepsilon_t \quad (3.114)$$

Les estimateurs  $\hat{\rho}$  et  $\hat{c}$  sont alors définis par le système matriciel :

$$\begin{bmatrix} \hat{c} \\ \hat{\rho} - 1 \end{bmatrix} = \begin{bmatrix} T & \sum_{t=1}^T x_{t-1} \\ \sum_{t=1}^T x_{t-1} & \sum_{t=1}^T x_{t-1}^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum_{t=1}^T \varepsilon_t \\ \sum_{t=1}^T x_{t-1} \varepsilon_t \end{bmatrix}$$

Etudions le comportement asymptotique de ces différents éléments :

$$\begin{aligned} T^{-1} \sum_{t=1}^T x_{t-1} \varepsilon_t &\xrightarrow[T \rightarrow \infty]{\mathcal{L}} \frac{\sigma_\varepsilon^2}{2} [W(1)^2 - 1] \\ T^{-2} \sum_{t=1}^T x_{t-1}^2 &\xrightarrow[T \rightarrow \infty]{\mathcal{L}} \sigma_\varepsilon^2 \int_0^1 [W(r)]^2 dr \\ T^{-\frac{3}{2}} \sum_{t=1}^T x_{t-1} &\xrightarrow[T \rightarrow \infty]{\mathcal{L}} \sigma_\varepsilon \int_0^1 W(r) dr \\ T^{-\frac{1}{2}} \sum_{t=1}^T \varepsilon_t &\xrightarrow[T \rightarrow \infty]{\mathcal{L}} \sigma_\varepsilon W(1) \end{aligned}$$

Afin de contrôler la vitesse de convergence des différents éléments de cette expression, on doit donc utiliser la transformée suivante<sup>11</sup> :

$$\begin{bmatrix} T^{\frac{1}{2}} \hat{c} \\ T(\hat{\rho} - 1) \end{bmatrix} = \begin{bmatrix} T^{-1} T & T^{-\frac{3}{2}} \sum_{t=1}^T x_{t-1} \\ T^{-\frac{3}{2}} \sum_{t=1}^T x_{t-1} & T^{-2} \sum_{t=1}^T x_{t-1}^2 \end{bmatrix}^{-1} \begin{bmatrix} T^{-\frac{1}{2}} \sum_{t=1}^T \varepsilon_t \\ T^{-1} \sum_{t=1}^T x_{t-1} \varepsilon_t \end{bmatrix} \quad (3.115)$$

Dès lors, en utilisant les propriétés de la convergence en loi, on montre que :

$$\begin{bmatrix} T^{\frac{1}{2}} \hat{c} \\ T(\hat{\rho} - 1) \end{bmatrix} \xrightarrow[T \rightarrow \infty]{\mathcal{L}} \begin{bmatrix} 1 & \sigma_\varepsilon \int_0^1 W(r) dr \\ \sigma_\varepsilon \int_0^1 W(r) dr & \sigma_\varepsilon^2 \int_0^1 [W(r)]^2 dr \end{bmatrix}^{-1} \begin{bmatrix} \sigma_\varepsilon W(1) \\ \frac{\sigma_\varepsilon^2}{2} [W(1)^2 - 1] \end{bmatrix}$$

A partir de ces différents éléments, on peut facilement dériver la distribution asymptotique de l'estimateur  $\hat{\rho}$ .

<sup>11</sup> En effet,  $\forall T \in \mathbb{R}^{+*}$ , et  $\forall (a, b, c, d, e, f) \in \mathbb{R}^6$  tels que  $ad \neq c^2$ , on a :

$$\begin{bmatrix} T^{-1}a & T^{-\frac{3}{2}}c \\ T^{-\frac{3}{2}}c & T^{-2}d \end{bmatrix}^{-1} \begin{bmatrix} T^{-\frac{1}{2}}e \\ T^{-1}f \end{bmatrix} = \begin{bmatrix} \sqrt{T} \frac{de - cf}{ad - c^2} \\ T \frac{-ce + af}{ad - c^2} \end{bmatrix}$$

**Proposition 21** *La distribution asymptotique de l'estimateur  $\hat{\rho}$  obtenue dans le modèle*

$$x_t = \rho x_{t-1} + c + \varepsilon_t \quad (3.116)$$

*sous l'hypothèse de non stationnarité  $H_0$ , est la suivante :*

$$T(\hat{\rho} - 1) \xrightarrow[T \rightarrow \infty]{\mathcal{L}} \frac{\frac{1}{2} [W(1)^2 - 1] - W(1) \int_0^1 W(r) dr}{\int_0^1 [W(r)]^2 dr - \left[ \int_0^1 W(r) dr \right]^2} \quad (3.117)$$

*Dans le modèle :*

$$x_t = \rho x_{t-1} + \varepsilon_t \quad (3.118)$$

*cette même distribution asymptotique de  $\hat{\rho}$  était définie par :*

$$T(\hat{\rho} - 1) \xrightarrow[T \rightarrow \infty]{\mathcal{L}} \frac{1}{2} \frac{[W(1)^2 - 1]}{\int_0^1 [W(r)]^2 dr} \quad (3.119)$$

On vérifie bien que la distribution de l'estimateur  $\hat{\rho}$  diffère suivant que l'on inclut ou non dans le modèle estimé une constante. Il en va de même pour l'introduction d'une tendance déterministe. Naturellement, si la distribution de  $\hat{\rho}$  diffère suivant les modèles, celle de la statistique de Student du test de Dickey Fuller diffère elle aussi suivant le modèle estimé. En particulier, on peut montrer (exercice) :

**Proposition 22** *Sous l'hypothèse  $H_0$  de non stationnarité, la distribution asymptotique de la statistique de Student associée au test de Dickey Fuller,  $\rho = 1$ , dans le modèle*

$$x_t = \rho x_{t-1} + c + \varepsilon_t \quad (3.120)$$

*avec  $\varepsilon_t$  i.i.d.  $(0, \sigma_\varepsilon^2)$ , est la suivante :*

$$t_{\hat{\rho}=1} \xrightarrow[T \rightarrow \infty]{\mathcal{L}} \frac{1}{2} \frac{[W(1)^2 - 1] - W(1) \int_0^1 W(r) dr}{\left\{ \int_0^1 [W(r)]^2 dr - \left[ \int_0^1 W(r) dr \right]^2 \right\}^{\frac{1}{2}}} \quad (3.121)$$

Ainsi, on vérifie que la présence d'une constante modifie la distribution asymptotique de la statistique de test de l'hypothèse de racine unitaire. En effet, dans le modèle sans constante, nous avons montré que :

$$t_{\hat{\rho}=1} \xrightarrow[T \rightarrow \infty]{\mathcal{L}} \frac{1}{2} \frac{[W(1)^2 - 1]}{\left\{ \int_0^1 [W(r)]^2 dr \right\}^{\frac{1}{2}}} \quad (3.122)$$

Il en est exactement de même concernant la présence d'une tendance dans le modèle. A chaque fois la distribution asymptotique de  $t_{\hat{\rho}=1}$  est affectée.

Tout ceci est d'autant plus regrettable pour l'économètre que les seuils critiques sont relativement différents pour les 3 modèles. Comme on peut l'observer sur le tableau (3.4), la déformation de la distribution asymptotique est suffisamment importante pour que les seuils standard soient très différents. On note  $C_{(\alpha)}^1$  le seuil critique pour un risque de première espèce de  $\alpha\%$  associé au modèle (3.110) sans constante ni trend. On note  $C_{(\alpha)}^2$  le seuil critique associé au modèle (3.111) avec constante sans trend. On note  $C_{(\alpha)}^3$  le seuil critique associé au modèle (3.112) avec constante et trend.

Tableau 3.4: Seuils Critiques des Tests  $t_{\hat{\phi}=0}$ 

	$\alpha = 1\%$	$\alpha = 5\%$	$\alpha = 10\%$
$C_{(\alpha)}^1$	-2.58	-1.95	-1.62
$C_{(\alpha)}^2$	-3.43	-2.86	-2.57
$C_{(\alpha)}^3$	-3.96	-3.41	-3.12

Les seuils sont tirés de Fuller (1976) pour  $T = \infty$

On comprend ainsi que si l'on se "*trompe*" de modèle, par exemple si l'on inclut à tort une constante dans le modèle estimé, on se "*trompe*" de seuil critique, et donc risque fort de se tromper de diagnostic quant à la stationnarité de la série étudiée. C'est pourquoi, il convient à présent de développer *une stratégie de tests de non stationnarité* pour tester la non stationnarité dans un modèle adéquat.

### 3.3 Une Stratégie de Tests

Nous allons à présent proposer *une stratégie de tests de Dickey Fuller permettant de tester la non stationnarité conditionnellement à la spécification du modèle utilisé*. On considère trois modèles définis comme suit :

$$\text{Modèle 1 : } \Delta x_t = \phi x_{t-1} + \varepsilon_t \quad (3.123)$$

$$\text{Modèle 2 : } \Delta x_t = \phi x_{t-1} + c + \varepsilon_t \quad (3.124)$$

$$\text{Modèle 3 : } \Delta x_t = \phi x_{t-1} + c + \beta t + \varepsilon_t \quad (3.125)$$

avec  $\varepsilon_t$  *i.i.d.*  $(0, \sigma_\varepsilon^2)$ . On cherche à tester l'hypothèse de racine unitaire :

$$H_0 : \phi = 0 \quad H_1 : \phi < 0 \quad (3.126)$$

**Remarque 1.** Le principe général de la stratégie de tests est le suivant. Il s'agit de partir du modèle le plus général, d'appliquer le test de racine unitaire en utilisant les seuils correspondant à ce modèle, puis de vérifier par un test approprié que le modèle retenu était le "bon". En effet, si le modèle n'était pas le "bon", les seuils utilisés pour le test de racine unitaire ne sont pas valables. On risque alors de commettre une erreur de diagnostic quant à la stationnarité de la série. Il convient dans ce cas, de recommencer le test de racine unitaire dans un autre modèle, plus contraint. Et ainsi de suite, jusqu'à trouver le "bon" modèle, les "bons" seuils et bien entendu les "bons" résultats.

Le déroulement de la stratégie de test est reportée sur la figure suivante. On commence par tester la racine unitaire à partir du modèle le plus général, à savoir le modèle 3. On compare la réalisation de la statistique de Student  $t_{\hat{\phi}=0}$  aux seuils  $C_{(\alpha)}^3$  tabulés par Dickey et Fuller, ou McKinnon pour le modèle 3 (par exemple  $-3.41$  à 5%, pour  $T \rightarrow \infty$ ). Si la réalisation de  $t_{\hat{\phi}=0}$  est supérieure<sup>12</sup> au seuil  $C_{(\alpha)}^3$ , on accepte l'hypothèse nulle de non stationnarité. Une fois que le diagnostic est établi, on cherche à vérifier si la spécification du modèle 3, incluant une constante et un trend, était une spécification compatible avec les données. On teste alors la nullité du coefficient  $\beta$  de la tendance. Deux choses l'une :

- Soit on a rejeté au préalable l'hypothèse de racine unitaire, dans ce cas on teste la nullité de  $\beta$  par un simple test de Student avec des seuils standards (test symétrique, donc seuil de 1.96 à 5%). Si l'on rejette l'hypothèse  $\beta = 0$ , cela signifie que le modèle 3 est le "bon" modèle

<sup>12</sup>Etant donné que le test est non symétrique, on ne considère pas, bien entendu, la valeur absolue de  $t_{\hat{\phi}=0}$ , mais son niveau relatif.

pour tester la racine unitaire, puisque la présence d'une tendance n'est pas rejetée. Dans ce cas, on conclut que la racine unitaire est rejetée, la série est *TS*, du fait de la présence de la tendance. En revanche, si l'on accepte l'hypothèse  $\beta = 0$ , le modèle n'est pas adapté puisque la présence d'une tendance est rejetée. On doit refaire le test de racine unitaire à partir du modèle 2, qui ne comprend qu'une constante.

- Soit, au contraire, on avait au préalable, accepté l'hypothèse de racine unitaire, et dans ce cas, on doit construire un test de Fischer de l'hypothèse jointe  $\phi = 0$  et  $\beta = 0$ . On teste ainsi la nullité de la tendance, conditionnellement à la présence d'une racine unitaire:

$$H_0^3 : (c; b; \phi) = (c; 0; 0) \text{ contre } H_1^3 \quad (3.127)$$

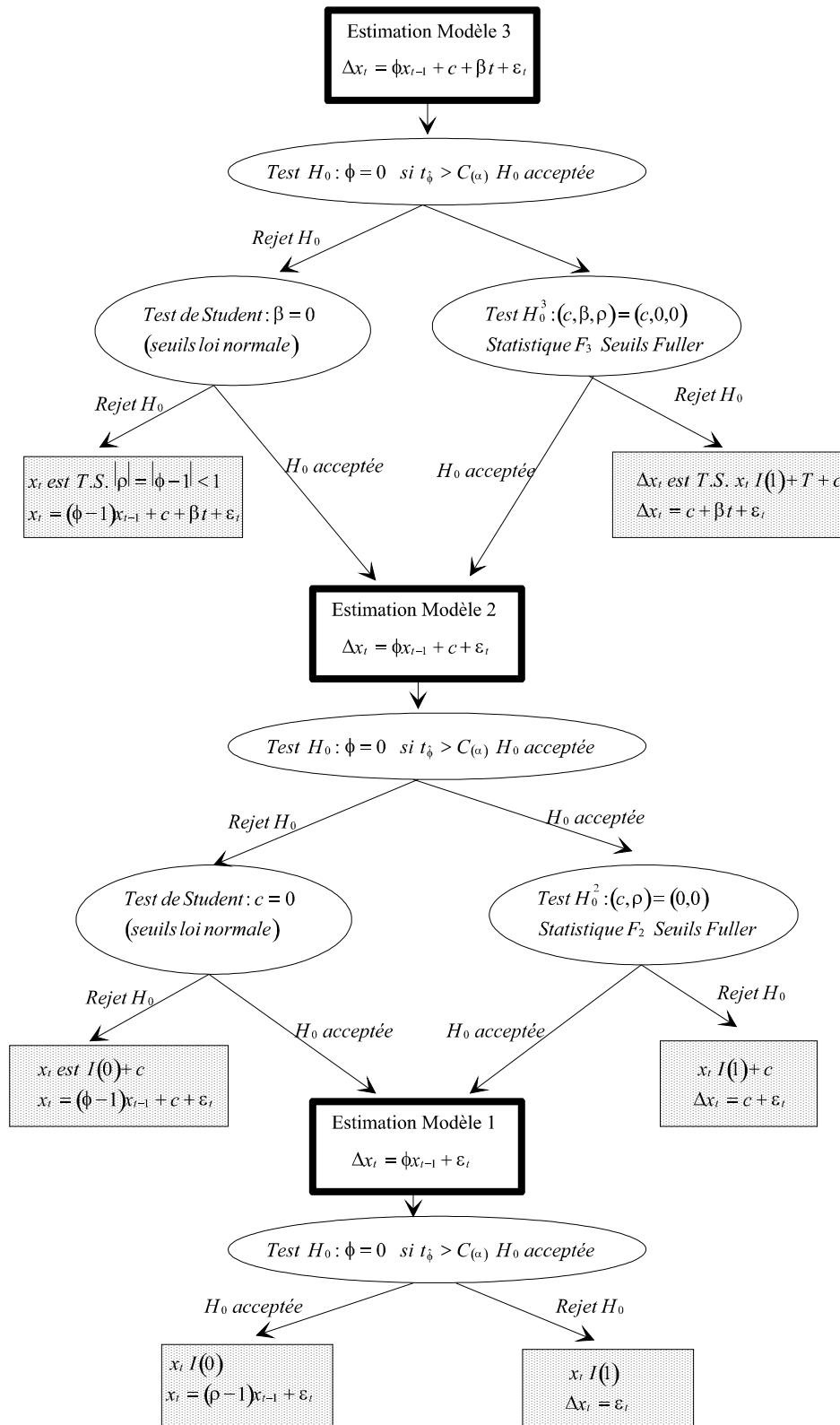
La statistique de ce test se construit de façon standard par la relation :

$$F_3 = \frac{(SCR_{3,c} - SCR_3) / 2}{SCR_3 / (T - 3)} \quad (3.128)$$

où  $SCR_{3,c}$  est la somme des carrés des résidus du modèle 3 contraint sous  $H_0^3$  :

$$\Delta x_t = c + \varepsilon_t$$

et  $SCR_3$  est la somme des carrés des résidus du modèle 3 non contraint (équation 3.125). Les seuils à distance fini de cette statistique sont fournies dans la partie annexe du programme de cours. Si la réalisation de  $F_3$  est supérieure à la valeur  $\phi_3$  lue dans la table à un seuil  $\alpha\%$ , on rejette l'hypothèse  $H_0^3$ . Dans ce cas, le modèle 3 est le "bon" modèle et la série  $x_t$  est intégrée d'ordre 1,  $I(1) + c + T$ , le taux de croissance est *TS*,  $\Delta x_t = c + \beta t + \varepsilon_t$ . En revanche, si l'on accepte  $H_0^3$ , le coefficient de la tendance est nul, le modèle 3 n'est pas le "bon" modèle, on doit donc effectuer à nouveau le test de non stationnarité dans le modèle 2.



Stratégie de Tests de Dickey Fuller



Si l'on a accepté la nullité du coefficient  $\beta$  de la tendance, on doit alors effectuer à nouveau les tests de non stationnarité à partir cette fois-ci du modèle 2 (équation 3.124) incluant uniquement une constante. On compare alors la réalisation de la statistique de Student  $t_{\hat{\phi}=0}$  aux seuils  $C_{(\alpha)}^2$  tabulés par Dickey et Fuller, ou McKinnon pour le modèle 2 (par exemple  $-2.86$  à 5%, pour  $T \rightarrow \infty$ ). Si la réalisation de  $t_{\hat{\phi}=0}$  est supérieure au seuil  $C_{(\alpha)}^2$ , on accepte l'hypothèse nulle de non stationnarité. Une fois que le diagnostic est établi, on cherche à vérifier si la spécification du modèle 2, incluant une constante, est une spécification compatible avec les données. On teste alors la nullité du coefficient  $c$  de la constante. Deux choses l'une :

- Soit on a rejeté au préalable l'hypothèse de racine unitaire, dans ce cas on teste la nullité de  $c$  par un simple test de Student avec des seuils standard (test symétrique, donc seuil de 1.96 à 5%). Si l'on rejette l'hypothèse  $c = 0$ , cela signifie que le modèle 2 est le "bon" modèle pour tester la racine unitaire, puisque la présence d'une constante n'est pas rejetée. Dans ce cas, on conclut que la racine unitaire est rejetée, la série est stationnaire  $I(0) + c$ . En revanche, si l'on accepte l'hypothèse  $c = 0$ , le modèle 2 n'est pas adapté puisque la présence d'une constante est rejetée. On doit refaire le test de racine unitaire à partir du modèle 1, qui ne comprend ni constante ni trend.
- Soit, au contraire, on avait au préalable, accepté l'hypothèse de racine unitaire, et dans ce cas, on doit construire un test de Fischer de l'hypothèse jointe  $\phi = 0$  et  $c = 0$ . On teste ainsi la nullité de la constante, conditionnellement à la présence d'une racine unitaire:

$$H_0^2 : (c; \phi) = (0; 0) \text{ contre } H_1^2 \quad (3.129)$$

La statistique de ce test se construit de façon standard par la relation :

$$F_2 = \frac{(SCR_{2,c} - SCR_2) / 2}{SCR_2 / (T - 2)} \quad (3.130)$$

où  $SCR_{2,c}$  est la somme des carrés des résidus du modèle 2 contraint sous  $H_0^2$ , c'est à dire  $SCR_{2,c} = \sum_{t=1}^T \varepsilon_t^2 = \sum_{t=1}^T (\Delta x_t)^2$  et  $SCR_2$  est la somme des carrés des résidus du modèle 2 non contraint (équation 3.124). Les seuils à distance fini de cette statistique sont fournies dans la partie annexe du programme de cours. Si la réalisation de  $F_2$  est supérieure à la valeur  $\phi_1$  lue dans la table à un seuil  $\alpha$ , on rejette l'hypothèse  $H_0^2$  au seuil  $\alpha\%$ . Dans ce cas, le modèle 2 est le "bon" modèle et la série  $x_t$  est intégrée d'ordre 1,  $I(1) + c$ . En revanche, si l'on accepte  $H_0^2$ , le coefficient de la constante est nul, le modèle 2 n'est pas le "bon" modèle, on doit donc effectuer à nouveau le test de non stationnarité dans le modèle 1.

Enfin, si l'on a accepté la nullité du coefficient  $c$  de la constante, on doit alors effectuer à nouveau les tests de non stationnarité à partir cette fois-ci du modèle 1 (équation 3.123) sans constante ni trend. On compare alors la réalisation de la statistique de Student  $t_{\hat{\phi}=0}$  aux seuils

$C_{(\alpha)}^1$  tabulés par Dickey et Fuller, ou McKinnon pour le modèle 1 (par exemple  $-1.95$  à  $5\%$ , pour  $T \rightarrow \infty$ ). Si la réalisation de  $t_{\hat{\phi}=0}$  est supérieure au seuil  $C_{(\alpha)}^1$ , on accepte l'hypothèse nulle de non stationnarité. Dans ce cas la série  $x_t$  est  $I(1)$  et correspond à une pure marche aléatoire,  $x_t = x_{t-1} + \varepsilon_t$ . Si l'hypothèse nulle est rejetée, la série est stationnaire,  $I(0)$  de moyenne nulle  $x_t = \rho x_{t-1} + \varepsilon_t$ , avec  $|\rho| < 1$ .

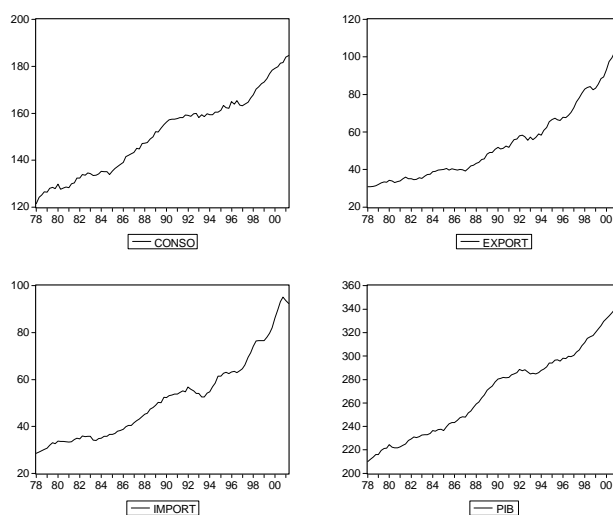
### 3.4 Application : PIB, consommation et commerce extérieur

Nous allons proposer à présent *une application de la stratégie de tests de Dickey Fuller* à différentes séries<sup>13</sup> issues des Comptes Nationaux Trimestriels de l'INSEE, relevées sur la période allant du première trimestre 1978 au deuxième trimestre 2001, soit 94 observations. Les quatre séries utilisées sont exprimées en milliards d'euros, au prix de 1995, et correspondent ainsi à des agrégats en volume qui sont de plus corrigés des variations saisonnières (*données CVS*).

1. CONSO : correspond à la consommation des ménages.
2. EXPORT : correspond aux exportations CAF agrégées.
3. IMPORT : correspond aux importations CAF agrégées.
4. PIB : correspond au produit intérieur brut.

Le graphique de ces séries est reporté sur la figure (3.11).

Figure 3.11: Comptes Trimestriels de l'INSEE : 1978:1-2001:2



<sup>13</sup>Ces séries sont disponibles sur le site de l'INSEE ([www.insee.fr](http://www.insee.fr)) dans la rubrique comptes nationaux.

Un simple examen graphique met clairement en évidence le fait que les quatre séries étudiées sont a priori non stationnaires. Les processus générateurs correspondants ne semblent pas satisfaire en effet la condition d'invariance de l'espérance, et il en va de même pour la variance. Reste à savoir si ces processus sont des processus *DS* ou *TS* selon la terminologie de Nelson et Plosser (1982). Nous faisons l'hypothèse<sup>14</sup> pour l'instant que les quatre processus, s'ils sont  $I(d)$ , sont au plus  $I(1)$ . Appliquons la stratégie de tests de Dickey Fuller exposée précédemment tout d'abord à la consommation des ménages.

On commence par estimer le modèle 3,  $\Delta x_t = \phi x_{t-1} + c + \beta.t + \varepsilon_t$ , incluant une constante et un trend. On teste alors la présence d'une racine unitaire dans le processus en testant la nullité du paramètre  $\phi$  à l'aide d'une statistique de Student  $t_{\hat{\phi}}$ , où  $\hat{\phi}$  désigne l'estimateur des MCO. Sous le logiciel Eviews, on peut soit effectuer directement cette régression en créant au préalable la série  $\Delta x_t$  et la tendance grâce à l'instruction @TREND(N), soit on utilise le test pré-programmé sous le logiciel. Pour cela, il suffit de cliquer sur la série, puis sur l'onglet *View*, puis sur l'onglet *Unit Root Test*. Apparaît alors une boîte de dialogue dans lequel on choisit le type *Augmented Dickey Fuller*, le *test in level*, on inclut une constante et un trend (*Trend and Intercept*) et l'on choisit un nombre de termes en différences retardés (*Lagged Difference*) égal à 0. Le résultat de l'affichage pour la série CONSO est reproduit sur la figure (3.12).

Figure 3.12: Test de Racine Unitaire sur CONSO : Modèle 3

ADF Test Statistic	-1.236344	1% Critical Value*	-4.0591
		5% Critical Value	-3.4581
		10% Critical Value	-3.1548

\*MacKinnon critical values for rejection of hypothesis of a unit root.

Augmented Dickey-Fuller Test Equation  
 Dependent Variable: D(CONSO)  
 Method: Least Squares  
 Date: 09/22/01 Time: 15:10  
 Sample(adjusted): 1978:2 2001:2  
 Included observations: 93 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
CONSO(-1)	-0.048939	0.039584	-1.236344	0.2195
C	6.490766	4.824305	1.345430	0.1819
@TREND(1978:1)	0.032860	0.024357	1.349076	0.1807

R-squared	0.023622	Mean dependent var	0.679151
Adjusted R-squared	0.001924	S.D. dependent var	1.007585
S.E. of regression	1.006615	Akaike info criterion	2.882791
Sum squared resid	91.19471	Schwarz criterion	2.964487
Log likelihood	-131.0498	F-statistic	1.088697
Durbin-Watson stat	2.208681	Prob(F-statistic)	0.341047

<sup>14</sup>Si l'on désire vérifier cette hypothèse, il suffit de tester la stationnarité des taux de croissance des séries en appliquant les tests de Dickey Fuller à ces taux de croissance.

Cette procédure nous donne la valeur des réalisations des estimateurs des MCO des différents paramètres du modèle 3 ( $\phi, c, \beta$ ) ainsi que les statistiques de Student associées aux tests de nullité de ces paramètres. Ce qui nous intéresse ici plus particulièrement c'est bien entendu la statistique de Student  $t_{\hat{\phi}}$  associée à la variable endogène retardée CONSO(-1). Celle-ci est ici égale à  $-1.236$ , et cette valeur est en outre reportée en haut de l'affichage (*ADF Test Statistic*). Pour tester l'hypothèse  $\phi = 0$ , on utilise alors les seuils tabulés par Dickey et Fuller (*cf.* document joint, photocopié d'exercices) pour le modèle 3 et pour une taille d'échantillon de 93 observations ( $T - 1$  en raison de la valeur retardée sur CONSO). Ces seuils sont reportés dans l'affichage de la procédure d'Eviews. Au seuil de 5%, le seuil critique est  $C_{(\alpha)} = -3.4581$  ( $-3.45$  dans la table fournie en annexe pour 100 observations). Ainsi, dans ce cas pour un niveau de risque de 5%,  $t_{\hat{\phi}} > C_{(\alpha)}$ , on accepte l'hypothèse nulle de racine unitaire ( $\phi = 0$ ).

Il faut à présent évaluer la validité de notre diagnostic en vérifiant que le modèle à partir duquel nous avons fait le test (modèle 3) est bien le "bon" modèle. Il nous faut donc à présent tester la nullité du coefficient de la tendance conditionnellement à la présence d'une racine unitaire. On effectue pour cela le test  $H_0^3 : (c, \beta, \phi) = (c, 0, 0)$ . Malheureusement, ce test joint n'est pas préprogrammé dans Eviews, il faut donc écrire le petit programme suivant pour obtenir la valeur de la statistique de Fisher  $F_3$  associée à  $H_0^3$ .

```

'- CONSTRUCTION DES DIFFERENCES PREMIERES
SMPL 1978:2 2001:2
GENR DCONSO = CONSO-CONSO(-1)
'- ESTIMATION DU MODELE LIBRE
EQUATION MOD3.LS DCONSO C @TREND(1978:1) CONSO(-1)
SCALAR SCR3=@SSR
SCALAR NDL=@REGOBS-@NCOEF
'- ESTIMATION DU MODELE CONTRAINT
EQUATION MOD3C.LS DCONSO C
SCALAR SCR3C=@SSR
'- CONSTRUCTION DE LA STATISTIQUE F3
SCALAR F3=((SCR3C-SCR3)/2)/(SCR3/NDL)

```

Ce programme n'appelle que peu de commentaires. Il utilise simplement l'instruction @SSR qui correspond à la somme des carrés des résidus de la régression précédente, l'instruction @REGOBS qui correspond au nombre d'observations de cette régression et l'instruction @NCOEF qui correspond au nombre de coefficients estimés. On estime<sup>15</sup> successivement le modèle libre (modèle 3) et le modèle

<sup>15</sup>Dans ce cas, bien entendu il n'est pas nécessaire d'appliquer les MCO pour obtenir la somme des carrés des résidus du modèle contraint puisque  $RC3_c = \sum_{t=1}^T (\Delta x_t - \overline{\Delta x_t})^2$  avec  $\overline{\Delta x_t} = (1/T) \sum_{t=1}^T \Delta x_t$ . Toutefois, nous verrons que dans le cas des tests *ADF*, il sera nécessaire de faire une telle régression. C'est pourquoi, nous présentons

contraint sous  $H_0^3$ , c'est à dire le modèle  $\Delta x_t = c + \varepsilon_t$ . La valeur de la réalisation de la statistique de Fischer  $F_3$  est stockée dans la variable scalaire (SCALAR) nommée F3. Il suffit de cliquer sur ce scalaire dans le Workfile d'EvIEWS pour voir s'afficher la valeur correspondante en bas de l'écran. Pour la variable de consommation, nous obtenons ainsi une valeur de  $F_3$  égale à 1.08. Cette valeur est à comparer aux seuils critiques lus dans la table de Dickey et Fuller (1981), tableau VI, page 1063, fournie en annexe (cf. polycopié d'exercices). Pour une taille d'échantillon de 100, et un risque de première espèce de 5%, la valeur critique est égale à 6.49. Donc la réalisation de  $F_3$  est inférieure au seuil critique, on *accepte l'hypothèse nulle de la nullité du coefficient de la tendance conditionnellement à la présence d'une racine unitaire*. Ceci signifie que le test de non stationnarité pratiqué avec les seuils asymptotiques incluant une tendance (modèle 3) doit être remis en cause. Il faut donc recommencer ce test à partir du modèle incluant uniquement une constante.

On estime alors le modèle 2,  $\Delta x_t = \phi x_{t-1} + c + \varepsilon_t$  et l'on teste la présence d'une racine unitaire. Pour cela, il suffit de cliquer sur la série, puis sur l'onglet *View*, puis sur l'onglet *Unit Root Test*. Apparaît alors une boîte de dialogue dans lequel on choisit le type *Augmented Dickey Fuller*, le *test in level*, on inclut uniquement une constante (*Intercept*) et l'on choisit un nombre de termes en différences retardés (*Lagged Difference*) égal à 0. Le résultat de l'affichage pour la série CONSO est reproduit sur la figure (3.13).

Figure 3.13: Test de Racine Unitaire sur CONSO : Modèle 2

ADF Test Statistic	0.595142	1% Critical Value*	-3.5015
		5% Critical Value	-2.8925
		10% Critical Value	-2.5831

\*MacKinnon critical values for rejection of hypothesis of a unit root.

Augmented Dickey-Fuller Test Equation				
Dependent Variable: D(CONSO)				
Method: Least Squares				
Date: 09/22/01 Time: 16:16				
Sample: 1978:2 2001:2				
Included observations: 93				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
CONSO(-1)	0.003778	0.006347	0.595142	0.5532
C	0.111345	0.959810	0.116008	0.9079
R-squared	0.003877	Mean dependent var		0.679151
Adjusted R-squared	-0.007069	S.D. dependent var		1.007585
S.E. of regression	1.011141	Akaike info criterion		2.881306
Sum squared resid	93.03888	Schwarz criterion		2.935770
Log likelihood	-131.9807	F-statistic		0.354194
Durbin-Watson stat	2.265070	Prob(F-statistic)		0.553225

le programme sous cette forme afin de présenter une démarche systématique.

La statistique de Student  $t_{\hat{\phi}}$  associée à la variable endogène retardée  $\text{CONSO}(-1)$  prend ici une valeur de 0.595. Pour tester l'hypothèse  $\phi = 0$ , on utilise alors les seuils tabulés par Dickey et Fuller pour le modèle 2 (*cf.* document joint, photocopié d'exercices) et pour une taille d'échantillon de 93 observations. Ces seuils sont reportés dans l'affichage de la procédure d'Eviews. Au seuil de 5%, le seuil critique est  $C_{(\alpha)} = -2.8925$  ( $-2.89$  dans la table fournie en annexe pour 100 observations). Ainsi, dans ce cas pour un niveau de risque de 5%,  $t_{\hat{\phi}} > C_{(\alpha)}$ , on accepte l'hypothèse nulle de racine unitaire ( $\phi = 0$ ) dans le modèle 2.

Il faut à nouveau évaluer la validité de notre diagnostic en vérifiant que le modèle 2 à partir duquel nous avons fait le test de racine unitaire est bien le "bon" modèle. On teste pour cela la nullité du coefficient de la constante conditionnellement à la présence d'une racine unitaire. On effectue le test  $H_0^2 : (c, \phi) = (0, 0)$ , ce qui revient, dans le cas du test de Dickey Fuller simple, à tester la nullité des deux coefficients du modèle 2. Dès lors, on peut utiliser la valeur de la réalisation de la statistique de Fisher programmé dans Eviews pour le test de la nullité de l'ensemble des coefficients du modèle. *Attention, ceci n'est valable que pour le test de Dickey Fuller simple et ne le sera plus pour les tests ADF que nous verrons ultérieurement.* Sur la figure (3.13), on observe que la réalisation de la statistique de Fisher, qui dans ce cas précis et uniquement dans ce cas précis correspond à  $F_2$ , est égale à 0.354. *Par contre, on ne doit surtout pas utiliser la pvalue programmé sous Eviews, puisque celle-ci est construite à partir d'une distribution de Fischer standard de  $F_2$ .* Il faut en effet comparer la réalisation de  $F_2$  aux seuils critiques de la table de Dickey et Fuller (1981), tableau VI, page 1063, fournie en annexe (*cf.* photocopié d'exercices). Pour une taille d'échantillon de 100, et un risque de première espèce de 5%, la valeur critique est égale à 4.71. Pour un risque de 5%, la réalisation de  $F_3$  est inférieure au seuil critique, on accepte l'hypothèse nulle de la nullité de la constante conditionnellement à la présence d'une racine unitaire. Ceci signifie que le test de non stationnarité pratiqué avec les seuils asymptotiques incluant une constante (modèle 2) doit être remis en cause. Il faut donc recommencer ce test à partir du modèle 1 sans constante, ni trend.

Il ne reste plus alors qu'à recommencer le test de racine unitaire à partir du modèle 1,  $\Delta x_t = \phi x_{t-1} + \varepsilon_t$ . Dans la boîte de dialogue, on choisit le type *Augmented Dickey Fuller*, le *test in level*, on exclut la constante et la tendance (*None*) et l'on choisit un nombre de termes en différences retardés (*Lagged Difference*) égal à 0. Le résultat de l'affichage pour la série  $\text{CONSO}$  est reproduit sur la figure (3.14). La statistique de Student  $t_{\hat{\phi}}$  associée à la variable endogène retardée prend ici une valeur de 6.538. Pour tester l'hypothèse  $\phi = 0$ , on utilise alors les seuils tabulés par Dickey et Fuller pour le modèle 1 (*cf.* document joint, photocopié d'exercices). Ces seuils sont reportés dans l'affichage de la procédure d'Eviews. Au seuil de 5%, le seuil critique est  $C_{(\alpha)} = -1.9436$  ( $-1.94$  dans la table fournie en annexe pour 100 observations). Ainsi, dans ce cas pour un niveau de risque de 5%,  $t_{\hat{\phi}} > C_{(\alpha)}$ , on accepte l'hypothèse nulle de racine unitaire ( $\phi = 0$ ) dans le modèle 1.

Figure 3.14: Test de Racine Unitaire sur CONSO : Modèle 1

ADF Test Statistic	6.538731	1% Critical Value*	-2.5800
		5% Critical Value	-1.9436
		10% Critical Value	-1.6175

\*MacKinnon critical values for rejection of hypothesis of a unit root.

Augmented Dickey-Fuller Test Equation

Dependent Variable: D(CONSO)

Method: Least Squares

Date: 09/22/01 Time: 22:28

Sample: 1978.2 2001.2

Included observations: 93

Variable	Coefficient	Std. Error	t-Statistic	Prob.
CONSO(-1)	0.004509	0.000690	6.538731	0.0000
R-squared	0.003730	Mean dependent var		0.679151
Adjusted R-squared	0.003730	S.D. dependent var		1.007585
S.E. of regression	1.005705	Akaike info criterion		2.859948
Sum squared resid	93.05263	Schwarz criterion		2.887180
Log likelihood	-131.9876	Durbin-Watson stat		2.286463

Finalement, selon le test de Dickey Fuller simple, *on conclut que la série trimestrielle de consommation des ménages est issue d'un processus non stationnaire, de type I(1) et peut être représentée par une pure marche aléatoire :*

$$x_t = x_{t-1} + \varepsilon_t \quad (3.131)$$

avec  $\varepsilon_t$  *i.i.d.*  $(0, \sigma_\varepsilon^2)$ . Pour stationnariser cette série, il convient donc de la différencier. En effet, on peut vérifier sur la figure (3.15) que la différence première de la série de consommation des ménages semble posséder un comportement de type stationnaire.

A l'issue de nos tests nous avons conclu que la différence première de la série correspond à l'innovation  $\varepsilon_t$ , qui dans le test de Dickey Fuller est assimilé à un bruit blanc. Or, il convient de s'assurer que la série différenciée possède bien les propriétés d'un bruit blanc. En particulier, il convient de s'assurer que celle-ci n'est pas autocorrélée puisque par définition  $E(\varepsilon_t \varepsilon_{t-k}) = 0$ , si  $k \neq 0$ . Pour cela, il nous faut étudier le corrélogramme de la série DCONSO. Pour obtenir ce corrélogramme sous Eviews, on clique sur la série, puis sur *Correlogram*. Pour un choix de lags de 10, les résultats affichés sont reportés sur la figure (3.16).

Pour un ordre  $k$  allant de 1 à 10, figurent ce corrélogramme la réalisation de l'autocorrélation empirique d'ordre  $k$  définie pour une série  $z_t$  par :

$$\hat{\rho}_k = \frac{1}{T-k} \sum_{t=1}^T [(z_t - \bar{z}_t)(z_{t-k} - \bar{z}_{t-k})] \quad (3.132)$$

Figure 3.15: Série de Consommation en Différences Premières

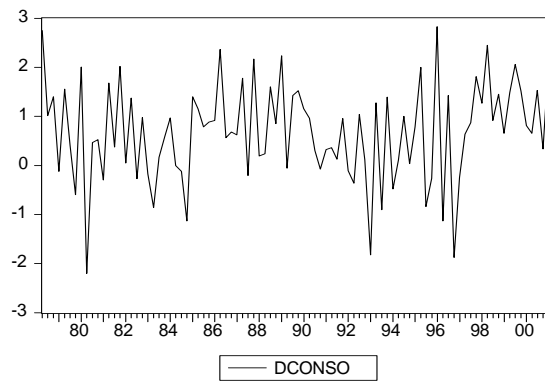


Figure 3.16: Corrélogramme de la Série DCONSO

Date: 09/22/01 Time: 22:41  
 Sample: 1978:2 2001:2  
 Included observations: 93

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
█	█	1 -0.157	-0.157	2.3560	0.125
█	█	2 0.274	0.256	9.6567	0.008
█	█	3 0.092	0.180	10.495	0.015
█	█	4 -0.030	-0.071	10.583	0.032
█	█	5 0.206	0.133	14.845	0.011
█	█	6 -0.079	-0.022	15.477	0.017
█	█	7 0.060	-0.044	15.847	0.027
█	█	8 -0.131	-0.164	17.640	0.024
█	█	9 0.128	0.131	19.360	0.022
█	█	10 -0.127	-0.065	21.085	0.021

On sait que  $\hat{\rho}_k$  converge en probabilité vers  $\rho_k = E[(z_t - \bar{z}_t)(z_{t-k} - \bar{z}_{t-k})]$ . Or, on observe dans la première colonne de la figure (3.16), notée AC pour *autocorrélation*, que l'autocorrélation de la série DCONSO notamment à l'ordre 2 est relativement importante. Elle est statistiquement différente de zéro puisque la réalisation sort de l'intervalle de la région de confiance de l'hypothèse de nullité matérialisée par des petits tirets verticaux. Cela signifie que la série de consommation différenciée est autocorrélée. Par conséquent, puisque nous avons conclu que  $\Delta x_t = \varepsilon_t$ , le processus  $\varepsilon_t$  n'est pas un bruit blanc. Or si le processus  $\varepsilon_t$  n'est pas un bruit blanc *i.i.d.*, cela remet en cause la validité de l'ensemble des distributions asymptotiques des statistiques de tests de Dickey Fuller et donc les conclusions que nous avons dressé quant à la non stationnarité de la série. *Il est donc nécessaire de tester la non stationnarité de la série en prenant en compte l'autocorrélation des perturbations  $\varepsilon_t$ . C'est précisément l'objet des tests de Dickey Fuller Augmentés, ou tests ADF.*



## 4 Tests de Dickey Fuller Augmentés

Comme nous l'avons vu dans le cadre de l'application sur la consommation des ménages, il arrive parfois que les résidus  $\varepsilon_t$  du modèle de Dickey Fuller soient autocorrélés. Or, les distributions asymptotiques des statistiques de test de racine unitaire ont été construites sous l'hypothèse que  $\varepsilon_t$  est un bruit blanc. Dès que l'on lève cette hypothèse les statistiques des tests de Dickey Fuller ne suivent plus les mêmes distributions asymptotiques, et donc les seuils de significativité des tests de racine unitaire sont différents. Il apparaît donc nécessaire de tenir compte de l'éventuelle autocorrélation des résidus dans la construction des tests de racine unitaire.

Il existe alors deux approches différentes pour tenir de cette éventuelle autocorrélation. La première approche, proposée par Phillips (1987) et Phillips et Perron (1988) consiste à proposer une correction des estimateurs des MCO et des statistiques de Student associées à ces estimateurs prenant en compte la possible autocorrélation des résidus. La seconde approche, développée par Dickey et Fuller (1979), consiste à contrôler directement l'autocorrélation dans le modèle (et non au niveau des estimateurs) en incluant un ou plusieurs termes autorégressifs différenciés. Nous allons montrer qu'une telle approche permet en effet de "blanchir" les résidus et de plus, de se ramener à une représentation similaire à celle du test de Dickey Fuller Simple. Dès lors, l'application de cette nouvelle stratégie est identique à celle présentée précédemment et l'on retrouve les mêmes distributions asymptotiques.

### 4.1 Les tests ADF : la prise en compte de l'autocorrélation des résidus

Pour bien comprendre l'argument de Dickey Fuller, considérons un processus  $x_t$  admettant une représentation de type  $AR(1)$  dont les innovations  $\varepsilon_t$  sont autocorrélées d'ordre  $p - 1$ .

$$x_t = \phi x_{t-1} + \varepsilon_t \quad (4.133)$$

où  $\varepsilon_t$  ne satisfait pas les hypothèses d'un bruit blanc et est autocorrélé d'ordre  $p - 1$  :

$$\varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_{p-1} \varepsilon_{t-p+1} = \mu_t \quad (4.134)$$

avec  $\mu_t$  *i.i.d.*  $(0, \sigma_\mu^2)$ . En substituant  $\varepsilon_t$  par l'expression  $x_t - \phi x_{t-1}$ , l'équation (4.134) se réécrit sous la forme :

$$\begin{aligned} \mu_t &= (x_t - \phi x_{t-1}) + \theta_1 (x_{t-1} - \phi x_{t-2}) + \theta_2 (x_{t-2} - \phi x_{t-3}) \dots + \theta_{p-1} (x_{t-p+1} - \phi x_{t-p}) \\ &= x_t + (\theta_1 - \phi) x_{t-1} + (\theta_2 - \phi \theta_1) x_{t-2} + \dots + (\theta_{p-1} - \phi \theta_{p-2}) x_{t-p+1} - \phi \theta_{p-1} x_{t-p} \end{aligned} \quad (4.135)$$

Ainsi, la représentation de type  $AR(1)$  avec autocorrélation des innovations d'ordre  $p-1$  peut être transformée en une représentation  $AR(p)$  où les innovations sont des bruits blancs.

$$x_t = \psi_1 x_{t-1} + \psi_2 x_{t-2} + \dots + \psi_p x_{t-p} + \mu_t \quad (4.136)$$

avec  $\psi_1 = -(\theta_1 - \phi)$ ,  $\psi_i = -(\theta_{i-1} - \phi\theta_{i-2})$  pour  $i = 2, \dots, p-1$  et  $\psi_p = \phi\theta_{p-1}$ .

**Remarque 1.** L'intuition de la démarche du test de Dickey Fuller Augmenté consiste à postuler un modèle de type  $AR(p)$  afin de corriger une éventuelle autocorrélation d'ordre  $p-1$  des innovations d'une représentation de type  $AR(1)$ .

Toutefois, le test de Dickey Fuller Augmenté n'est pas construit directement à partir de la forme  $AR(p)$  avec innovations *i.i.d.* de l'équation (4.135). Les auteurs privilégient une spécification incluant  $p-1$  termes différenciés retardés, connue sous le nom de représentation de Sims, Stock et Watson (1990).

**Proposition 23** (Représentation de Sims, Stock et Watson 1990) **Tout processus**  $(x_t, t \in \mathbb{Z})$  **satisfaisant une représentation**  $AR(p)$

$$x_t = \psi_1 x_{t-1} + \psi_2 x_{t-2} + \dots + \psi_p x_{t-p} + \mu_t \quad (4.137)$$

avec  $\mu_t$  *i.i.d.*  $(0, \sigma_\mu^2)$ , peut être exprimé sous la forme suivante :

$$x_t = \rho x_{t-1} + \xi_1 \Delta x_{t-1} + \xi_2 \Delta x_{t-2} + \dots + \xi_{p-1} \Delta x_{t-(p-1)} + \mu_t \quad (4.138)$$

où  $\xi_j$  désigne l'opposé de la somme partielle des coefficients  $\forall j = 0, 1, \dots, p-1$  :

$$\xi_j = -[\psi_{j+1} + \psi_{j+2} + \dots + \psi_p] \quad (4.139)$$

$$\rho = -\xi_0 = \psi_1 + \psi_2 + \dots + \psi_p \quad (4.140)$$

Cette forme canonique est connue sous le nom de représentation de Sims, Stock et Watson (1990).

La démonstration de cette proposition est la suivante. On considère un processus  $AR(p)$  :

$$\Psi(L)x_t = (1 - \psi_1 L - \psi_2 L^2 - \dots - \psi_p L^p)x_t = \mu_t \quad (4.141)$$

avec  $\mu_t$  *i.i.d.*  $(0, \sigma_\mu^2)$ . Notons que cette équation est formellement identique à l'équation (4.135). Nous allons à présent exprimer ce modèle en introduisant des termes différenciés retardés. On commence par poser :

$$\begin{aligned} & 1 - \psi_1 L - \psi_2 L^2 - \dots - \psi_{p-1} L^{p-1} - \psi_p L^p \\ = & 1 - \psi_1 L - \psi_2 L^2 - \dots - \psi_{p-2} L^{p-2} - [-(\psi_p) + (\psi_{p-1} + \psi_{p-1})] L^{p-1} - \psi_p L^p \\ = & 1 - \psi_1 L - \psi_2 L^2 - \dots - \psi_{p-3} L^{p-3} - [-(\psi_p + \psi_{p-1}) + (\psi_{p-2} + \psi_{p-1} + \psi_p)] L^{p-2} \\ & - [-(\psi_p) + (\psi_{p-1} + \psi_p)] L^{p-1} - \psi_p L^p \end{aligned}$$

En poursuivant cette démarche, on obtient une expression du polynôme  $\Psi(L)$  faisant apparaître les sommes partielles des coefficients, notées  $\xi_j$  :

$$\begin{aligned} \Psi(L) &= 1 - (\xi_1 - \xi_0)L - (\xi_2 - \xi_1)L^2 - (\xi_3 - \xi_2)L^3 \\ &\quad \dots - (\xi_{p-1} - \xi_{p-2})L^{p-1} - (-\xi_{p-1})L^p \end{aligned} \quad (4.142)$$

avec

$$\xi_j = -[\psi_{j+1} + \psi_{j+2} + \dots + \psi_p] \quad j = 0, 1, \dots, p-1 \quad (4.143)$$

En arrangeant les termes du polynôme  $\Psi(L)$ , on parvient alors à la formulation suivante :

$$\Psi(L) = (1 + \xi_0 L) - (\xi_1 L + \xi_2 L^2 + \dots + \xi_{p-1} L^{p-1})(1 - L) \quad (4.144)$$

C'est cette expression de  $\Psi(L)$  qui nous permet de faire apparaître un polynôme retard défini en la différence première  $(1 - L)$ . En posant  $\rho = -\xi_0$ , on retrouve alors la représentation de Sims, Stock et Watson (équation 4.138).

Maintenant, envisageons le cas où le processus  $AR(p)$  admet au plus une racine unitaire. Si tel est le cas, le polynôme  $\Psi(L)$  admet un pour racine.

$$\Psi(1) = 0 \iff 1 - \psi_1 - \psi_2 - \dots - \psi_p = 0 \quad (4.145)$$

Sous cette hypothèse, le terme  $\rho = -\xi_0$  de la représentation de Sims, Stock et Watson est alors égal à l'unité.

**Proposition 24** *Si le polynôme  $\Psi(L)$  associé au processus  $(x_t, t \in \mathbb{Z})$ , admet une racine unitaire, alors le terme  $\rho$  dans la représentation de Sims, Stock et Watson (équation 4.138) est égal à l'unité. Le processus  $(x_t, t \in \mathbb{Z})$  se réécrit alors sous la forme :*

$$\Delta x_t = \xi_1 \Delta x_{t-1} + \xi_2 \Delta x_{t-2} + \dots + \xi_{p-1} \Delta x_{t-(p-1)} + \mu_t \quad (4.146)$$

*Dès lors, le test de l'hypothèse de racine unitaire dans un modèle ADF revient à tester la nullité du coefficient  $\phi$  dans le modèle incluant  $p-1$  termes différenciés retardés :*

$$\Delta x_t = \phi x_{t-1} + \xi_1 \Delta x_{t-1} + \xi_2 \Delta x_{t-2} + \dots + \xi_{p-1} \Delta x_{t-(p-1)} + \mu_t \quad (4.147)$$

*avec  $\phi = \rho - 1$ . Cette structure est identique à celle des tests de Dickey Fuller Simples aux termes retardés près.*

On comprend ainsi que l'idée de Dickey Fuller consiste à se ramener à une représentation similaire à celle du test de Dickey Fuller Simple, mais qui leur permet de traiter le problème de l'autocorrélation des innovations.

**Résumé de la démarche** Partant d'un modèle  $AR(1)$  (équation 4.133), on sait que les résidus  $\varepsilon_t$  sont autocorrélés d'ordre  $p-1$ . Donc on peut pas faire le test directement à partir de cette représentation. On cherche à se ramener à une représentation alternative dans laquelle les innovations sont des bruits blancs. Pour ce faire, on se ramène à un  $AR$  d'ordre  $p$  (équation 3.87). Mais si l'on désire tester la racine unitaire directement à partir de l'équation 3.87, il est nécessaire de proposer une stratégie de test différente de celle du test de Dickey Fuller Simple. Donc, les auteurs ont cherché une autre représentation de l' $AR(p)$  leur permettant de pratiquer un test de racine unitaire dont la spécification est exactement identique à celle du test de Dickey Fuller Simple : c'est la représentation de Sims, Stock et Watson (équation 4.138). Ne reste plus qu'à démontrer que les distributions asymptotiques des statistiques de tests sont identiques à celles des tests de Dickey Fuller Simples.

Afin de mieux comprendre cette démarche considérons l'exemple suivant.

**Exemple :** On considère le processus  $z_t$  satisfaisant une représentation  $AR(1)$  où les résidus  $\varepsilon_t$  sont autocorrélés :

$$\begin{aligned} z_t &= \frac{1}{2}z_{t-1} + \varepsilon_t \\ \varepsilon_t &= \varepsilon_{t-1} + \mu_t \end{aligned}$$

avec  $\mu_t$  i.i.d.  $(0, \sigma_\mu^2)$ . Dans ce cas, les résidus  $\varepsilon_t$  sont autocorrélés d'ordre 1, on peut donc trouver une représentation de type  $AR(2)$  sur  $z_t$  avec des innovations de type bruits blancs :

$$\varepsilon_t = \varepsilon_{t-1} + \mu_t \iff \left( z_t - \frac{1}{2}z_{t-1} \right) = \left( z_{t-1} - \frac{1}{2}z_{t-2} \right) + \mu_t \quad (4.148)$$

On obtient ainsi :

$$z_t - \frac{3}{2}z_{t-1} + \frac{1}{2}z_{t-2} = \Psi(L) z_t = \mu_t$$

A partir de cet  $AR(2)$ , on cherche la représentation canonique de Sims, Stock et Watson :

$$\begin{aligned} \Psi(L) &= 1 - \frac{3}{2}L + \frac{1}{2}L^2 \\ &= 1 + \left[ \left( -\frac{3}{2} + \frac{1}{2} \right) - \frac{1}{2} \right] L + \frac{1}{2}L^2 \\ &= 1 - L - \frac{1}{2}L + \frac{1}{2}L^2 \end{aligned}$$

On obtient ainsi la représentation suivante :

$$\Psi(L) = (1 + \xi_0 L) - (\xi_1 L)(1 - L) = (1 - L) - \left( \frac{1}{2}L \right) (1 - L)$$

avec  $\xi_0 = -(\psi_1 + \psi_2) = -\left(\frac{3}{2} - \frac{1}{2}\right) = -1$  et  $\xi_1 = -\psi_2 = \frac{1}{2}$ . Le modèle s'écrit donc en un polynôme de degré 1 défini en la différence première  $(1 - L) z_t$  :

$$\Psi(L) z_t = \Delta z_t - \frac{1}{2}\Delta z_{t-1} = \mu_t \quad (4.149)$$

ou encore :

$$z_t = z_{t-1} + \frac{1}{2}\Delta z_{t-1} + \mu_t \quad (4.150)$$

On a ici inclut 1 terme différencié retardé pour corriger l'autocorrélation d'ordre 1 des résidus du modèle initial. De plus, d'après la proposition précédente, on sait que le processus  $z_t$  comporte une racine unitaire et est par là même  $I(1)$ .

Ainsi, pour un choix de  $p$  retards, correspondant à une autocorrélation d'ordre  $p + 1$  des innovations dans une représentation  $AR(1)$ , les trois modèles utilisés pour développer le test  $ADF$  sont les suivants :

$$\text{Modèle 1 : } \Delta x_t = \phi x_{t-1} + \sum_{j=1}^p \xi_j \Delta x_{t-j} + \mu_t \quad (4.151)$$

$$\text{Modèle 2 : } \Delta x_t = \phi x_{t-1} + \sum_{j=1}^p \xi_j \Delta x_{t-j} + c + \mu_t \quad (4.152)$$

$$\text{Modèle 3 : } \Delta x_t = \phi x_{t-1} + \sum_{j=1}^p \xi_j \Delta x_{t-j} + c + \beta t + \mu_t \quad (4.153)$$

**Proposition 25** *La stratégie de test  $ADF$  consiste en un première étape à déterminer le nombre de retard  $p$  nécessaire pour blanchir les résidus. Dans la seconde étape, il suffit d'appliquer la stratégie séquentielle du test de Dickey Fuller Simple aux modèles (4.151), (4.152) et (4.153). Les distributions asymptotiques des statistiques de test  $t_{\hat{\phi}}$  obtenues dans ces trois modèles sont alors identiques à celles obtenues dans les modèles de Dickey Fuller Simple correspondants.*

La démonstration de cette proposition sera fournie à titre d'exercice. Pour l'économètre appliqué, cela implique que les seuils de significativité pour les tests de racine unitaire  $DF$  et  $ADF$  sont identiques, et c'est là le principale avantage de démarche de Dickey Fuller (1979). Ces deux auteurs ont ainsi fourni une démarche de tests intégrée, avec des seuils identiques en cas d'autocorrélation ou non des résidus, grâce à l'adoption de la représentation de Sims, Stock et Watson.

## 4.2 Choix du nombre de retards optimal

Jusqu'à présent nous avons supposé que l'ordre d'autocorrélation des résidus  $\varepsilon_t$  était connu. Si cet ordre est égal à  $p$ , on sait à présent qu'il faut inclure dans le test  $ADF$ ,  $p$  termes différenciés retardés, ce qui correspond alors à un modèle de type  $AR(p + 1)$ . Le problème, c'est que dans la "vraie" vie, on ignore souvent *a priori* l'ordre d'autocorrélation des résidus. On doit donc chercher le nombre optimal  $p$ . Pour ce faire, plusieurs approches peuvent être envisagées, parmi celles-ci nous n'en retiendrons que deux : le contrôle *ex-post* de l'absence d'autocorrélation des innovations ou la minimisation de critères d'information.

### 4.2.1 Critères d'information

Une des manières de choisir le nombre de retards consiste à comparer différents modèles *ADF* incluant différents choix de retards, sur la base de *critères d'information*. Un critère d'information est un critère fondé sur le pouvoir prédictif du modèle considéré et qui tient du nombre de paramètres à estimer. *De façon concrète ces critères sont construits comme des fonctions de la variance des résidus estimés du modèle  $\sigma_{\hat{\varepsilon}_t}^2$  et du nombre de paramètres à estimer. L'objectif étant bien entendu de minimiser cette fonction par rapport à ces deux arguments (application du principe de parcimonie).* Ces critères s'applique de façon générale à tout type de modèle et pas uniquement aux modèles des tests *ADF*. Nous en retiendrons que deux : *le critère d'Akaike* et *le critère de Schwarz* (1978).

**Definition 26** *Pour un modèle, incluant  $k$  paramètres, estimé sur  $T$  périodes et dont la réalisation de l'estimateur de la variance des résidus est  $\sigma_{\hat{\varepsilon}_t}^2$ , le critère d'Akaike, ou AIC, est :*

$$AIC(k) = T \log(\sigma_{\hat{\varepsilon}_t}^2) + 2(k) \quad (4.154)$$

*Le critère de Schwartz (1978) est défini par :*

$$SC(k) = T \log(\sigma_{\hat{\varepsilon}_t}^2) + k \log(T) \quad (4.155)$$

Dans le cas de notre application au choix du nombre de retards dans les modèles *ADF*, on cherche le nombre de retard  $p$  qui minimise ces deux critères. Prenons par exemple, le cas du modèle *ADF* 3 :

$$\Delta x_t = \phi x_{t-1} + \sum_{j=1}^p \xi_j \Delta x_{t-j} + c + \beta t + \mu_t \quad (4.156)$$

Dans ce cas, pour un choix de retards  $p$  on a un nombre de paramètres à estimer égal à  $k = 3 + p$ . Si l'on dispose d'un échantillon de taille  $T$  et que l'on note  $\sigma_{\hat{\varepsilon}_t}^2$  la réalisation de l'estimateur de la variance des résidus obtenue dans le modèle avec  $p$  retard, les deux critères sont définis en fonction de  $p$  par :

$$AIC(p) = T \log(\sigma_{\hat{\varepsilon}_t}^2) + 2(3 + p) \quad (4.157)$$

$$SC(p) = T \log(\sigma_{\hat{\varepsilon}_t}^2) + (3 + p) \log(T) \quad (4.158)$$

On cherche donc le nombre de retards  $p$  qui minimise ces deux critères. Sous Eviews, ces deux critères sont fournis dès lors que l'on utilise l'instruction de régression. Nous verrons dans l'application comment obtenir la réalisation de ces deux critères.

### 4.2.2 Tests d'autocorrélation des résidus

Une fois que l'on est parvenu à apporter un diagnostic quant à la non stationnarité de la série et à identifier un modèle final, on peut obtenir une réalisation du processus des innovations  $\{\varepsilon_t, t \in \mathbb{Z}\}$ . Il convient alors de vérifier que celle-ci satisfait les propriétés d'un bruit blanc. *Les tests d'autocorrélation de résidus ne peuvent donc être mis en place qu'ex-post : ils ne servent pas directement au choix ex-ante du nombre de retards, ils ne servent qu'à valider un choix de retards.* Par exemple, supposons que l'on ait identifié un modèle du type :

$$\Delta x_t = \phi x_{t-1} + \xi_1 \Delta x_{t-1} + c + \mu_t \quad (4.159)$$

avec  $\phi = 0, p = 1$ . Le processus  $x_t$  est donc  $I(1)$ . En estimant le modèle  $\Delta x_t = \xi_1 \Delta x_{t-1} + c + \mu_t$ , on obtient une réalisation de taille  $T$  des innovations  $\mu_t$ . On doit alors vérifier l'absence d'autocorrélation des  $\mu_t$ .

**Remarque 1.** Si il n'y pas d'autocorrélation, le modèle est bien spécifié. Si en revanche, il existe de l'autocorrélation, le modèle est mal spécifié et en particulier le choix  $p = 1$  n'est pas valide. Il convient d'augmenter le nombre de retards afin de corriger l'autocorrélation des résidus.

Pour tester l'autocorrélation, on étudie le corrélogramme de la série et l'on utilise le test dit du "porte-manteau"<sup>16</sup>, dont il existe deux variantes :

**Definition 27** On note  $r_n$  l'autocorrélation empirique d'ordre  $n$  des résidus  $\varepsilon_t$  d'un modèle incluant  $k$  paramètres et estimé sur  $T$  périodes. Pour un ordre  $N$ , le test de Box et Pierce est le test de l'hypothèse  $H_0 : r_1 = \dots = r_N = 0$  contre  $H_1 : \exists j \in [1, N]$ , tel que  $r_j \neq 0$ . La statistique de ce test est :

$$Q_{BP} = T \sum_{n=1}^N r_n^2 \xrightarrow[T \rightarrow \infty]{\mathcal{L}} \chi^2(N - k) \quad (4.160)$$

L'hypothèse  $H_0$  est rejetée au seuil de 5% si  $Q_{BP}$  est supérieur au quantile 0.95 de la loi du  $\chi^2$  correspondant.

**Definition 28** Le test du Ljung-Box, pour un ordre  $N$ , correspond à l'hypothèse nulle  $H_0 : r_n = 0 \forall n \leq N$  et sont construites de la façon suivante :

$$Q_K = T(T + 2) \sum_{n=1}^N \frac{r_n^2}{T - n} \xrightarrow[T \rightarrow \infty]{\mathcal{L}} \chi^2(N - k) \quad (4.161)$$

<sup>16</sup>Ou "fourre tout" dans une traduction plus précise du terme anglais.

Sous Eviews, l'instruction *Correlogram* permet d'obtenir les *Qstats* associées au second test, celui de *Ljung-Box*. On peut ainsi facilement tester l'autocorrélation des résidus empiriques d'un modèle *ADF* incluant  $p$  retards.

### 4.3 Application des tests ADF : la consommation des ménages

Nous avons montré précédemment la nécessité d'appliquer les tests *ADF* à la série de consommation des ménages issues des comptes nationaux trimestriels, en raison de la présence d'une autocorrélation des résidus du modèle retenu. C'est ce que nous allons faire à présent et nous déterminer la correction de l'autocorrélation modifie le diagnostic quant à la non stationnarité de la série. Rappelons que cette série est disponible du premier trimestre 1978 au deuxième trimestre 2001, soit sur 94 observations. La série est exprimée en milliards d'euros, au prix de 1995, et est corrigée des variations saisonnières (*données CVS*). Nous allons dans un premier temps chercher à déterminer l'ordre des retards optimal, puis nous appliquerons la stratégie de test *ADF* à la série CONSO.

#### 4.3.1 Choix optimal des retards

Comme nous l'avons indiqué précédemment, il existe différentes façons de choisir l'ordre optimal  $p^*$  des retards dans le modèle des tests Dickey Fuller Augmentés. Dans la pratique, on se limite souvent à l'observations des critères d'information et à la vérification *ex-post* de l'absence d'autocorrélation des innovations.

Tableau 4.5: Choix du Nombre de Retards

$p$	Modèle 3		Modèle 2		Modèle 1	
	AIC	SC	AIC	SC	AIC	SC
5	2.765	2.990	2.810	3.007	2.791	2.960
4	2.778	2.974	2.800	2.967	2.779	2.919
3	<b>2.754</b>	2.921	<b>2.782</b>	2.921	<b>2.764</b>	2.875
2	2.779	<b>2.917</b>	2.785	<b>2.895</b>	2.766	<b>2.849</b>
1	2.838	2.948	2.827	2.910	2.808	2.862
0	2.882	2.964	2.881	2.935	2.859	2.887

Dans un premier temps, on utilise ainsi les critères d'information disponibles sous Eviews, à savoir le *critère d'Akaike* (AIC) et le *critère de Schwartz* (tableau 4.5). Pour ce faire, il suffit tout d'abord de se donner un nombre de retards maximum admissibles, noté  $p_{\max}$ , compte tenu du nombre d'observations disponibles et donc du nombre de degrés de liberté des régressions correspondantes. Dans cette étude nous poserons  $p_{\max} = 5$ . Puis, pour chaque modèle, on cherche le nombre de retards  $p^*$  optimal, compris entre 0 (test de Dickey Fuller simple) et  $p_{\max}$  qui minimise les deux critères d'informations. Une des façons de d'obtenir les deux critères d'information sous



EvIEWS consiste à cliquer sur la série, sur *Unit Root Test*, de choisir le modèle (*Intercept and Trend*, *Intercept* ou *None*) et le nombre de lags, puis de recommencer l'opération pour  $p$  variant de 0 à  $p_{\max}$ . Les critères d'information (*Akaike Info Criterion* et *Schwarz criterion*) figurent alors sur l'écran d'affichage des résultats (voir par exemple la figure 3.14). Après avoir répété cette opération, les résultats obtenus ont été reportés dans le tableau 4.5.

Quel que soit le modèle retenu, on constate que le critère d'Akaike conduit à un choix de retard optimal  $p^* = 3$ , tandis que le critère de Schwartz conduit à  $p^* = 2$ . On est donc ici en présence d'une divergence de diagnostic quant à l'utilisation de ces deux critères d'information, ce qui arrive souvent dans la pratique. Dans ce cas, il est nécessaire de bien comprendre que *l'objectif de l'introduction des termes retardés consiste à blanchir les résidus, c'est à dire à contrôler l'autocorrélation des innovations*. Dès lors, on cherche la structure minimale qui permet d'atteindre cet objectif. *Selon un principe de parcimonie, il convient de choisir le modèle incluant le minimum de paramètres à estimer et qui permet de blanchir totalement les résidus*. On adopte donc ici un choix optimal de retard  $p^* = 2$  et nous vérifierons ex-post dans le modèle retenu (avec ou sans constante) que l'introduction des deux termes différenciés retardés a permis d'éliminer totalement l'autocorrélation des résidus.

### 4.3.2 Stratégie de tests ADF

Comme nous l'avons expliqué précédemment, la stratégie de test des Dickey Fuller Augmentés est strictement identique à celle des tests Dickey Fuller simple, mis à part la modification des 3 modèles qui inclut à présent des termes différenciés retardés.

On commence par estimer le modèle 3 incluant une constante, un trend et deux termes différenciés retardés :

$$\Delta x_t = \phi x_{t-1} + c + \beta.t + \sum_{i=1}^2 \gamma_i \Delta x_{t-i} + \mu_t \quad (4.162)$$

avec  $\mu_t$  *i.i.d.*  $(0, \sigma_\mu^2)$ . On teste alors la présence d'une racine unitaire dans le processus en testant la nullité du paramètre  $\phi$  à l'aide d'une statistique de Student  $t_{\hat{\phi}}$ , où  $\hat{\phi}$  désigne l'estimateur des MCO. Pour effectuer ce test sous le logiciel EvIEWS, on clique sur la série, puis sur l'onglet *View*, puis sur l'onglet *Unit Root Test*. Apparaît alors une boîte de dialogue dans lequel on choisit le type *Augmented Dickey Fuller*, le *test in level*, on inclut une constante et un trend (*Trend and Intercept*) et l'on choisit un nombre de termes en différences retardés (*Lagged Difference*) égal à 2. Le résultat de l'affichage pour la série CONSO est reproduit sur la figure (4.17).

Pour cet échantillon de 91 observations (après les ajustements dus aux valeurs retardées), la réalisation de la statistique de Student  $t_{\hat{\phi}}$  est égale à  $-1.344$ . On compare cette valeur aux seuils<sup>17</sup>

<sup>17</sup>On utilise ici les mêmes seuils que pour le test de Dickey Fuller Simple.

Figure 4.17: Test de Racine Unitaire *ADF* sur CONSO : Modèle 3

ADF Test Statistic	-1.344024	1% Critical Value*	-4.0613
		5% Critical Value	-3.4591
		10% Critical Value	-3.1554

\*MacKinnon critical values for rejection of hypothesis of a unit root.

Augmented Dickey-Fuller Test Equation

Dependent Variable: D(CONSO)

Method: Least Squares

Date: 09/24/01 Time: 10:25

Sample(adjusted): 1978:4 2001:2

Included observations: 91 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
CONSO(-1)	-0.052967	0.039409	-1.344024	0.1825
D(CONSO(-1))	-0.109090	0.107113	-1.018460	0.3113
D(CONSO(-2))	0.278524	0.103227	2.698183	0.0084
C	6.746635	4.778293	1.411934	0.1616
@TREND(1978:1)	0.037225	0.024044	1.548158	0.1253
R-squared	0.135064	Mean dependent var	0.652692	
Adjusted R-squared	0.094834	S.D. dependent var	0.994087	
S.E. of regression	0.945777	Akaike info criterion	2.779758	
Sum squared resid	76.92645	Schwarz criterion	2.917717	
Log likelihood	-121.4790	F-statistic	3.357326	
Durbin-Watson stat	2.137636	Prob(F-statistic)	0.013284	

tabulés par Dickey et Fuller (*cf.* document joint, photocopié d'exercices) pour le modèle 3 et pour une taille d'échantillon de 91 observations. Au seuil de 5%, le seuil critique est  $C_{(\alpha)} = -3.4591$  ( $-3.45$  dans la table fournie en annexe pour 100 observations). Ainsi, dans ce cas pour un niveau de risque de 5%,  $t_{\hat{\phi}} > C_{(\alpha)}$ , on accepte *l'hypothèse nulle de racine unitaire* ( $\phi = 0$ ).

On vérifie ensuite que le modèle à partir duquel nous avons fait le test (modèle 3) est bien le "bon" modèle. On teste pour cela la nullité du coefficient de la tendance conditionnellement à la présence d'une racine unitaire : c'est le test  $H_0^3$ ,  $(c, \beta, \phi) = (c, 0, 0)$ . Pour obtenir la réalisation de la statistique de Fisher  $F_3$  associée à  $H_0^3$ , on construit le même programme que celui exposé dans le cadre des tests *DF* simples. La seule différence réside dans le fait que dans les deux régressions (modèle libre et modèle contraint), on doit maintenant ajouter les termes  $DCONSO(-1)$  et  $DCONSO(-2)$ . Pour la variable de consommation, nous obtenons ainsi une valeur de  $F_3$  égale à 1.8725. Cette valeur est à comparer aux seuils critiques lus dans la table de Dickey et Fuller (1981), tableau VI, page 1063, fournie en annexe (*cf.* photocopié d'exercices). Pour une taille d'échantillon de 100, et un risque de première espèce de 5%, la valeur critique est égale à 6.49. Donc la réalisation de  $F_3$  est inférieure au seuil critique, on *accepte l'hypothèse nulle de la nullité du coefficient de la tendance conditionnellement à la présence d'une racine unitaire*. Ceci signifie que le test de non stationnarité pratiqué avec les seuils asymptotiques incluant une tendance (modèle 3) doit être remis en cause.

Il faut donc recommencer ce test à partir du modèle incluant uniquement une constante.

On estime à présent le modèle 2 et l'on teste la présence d'une racine unitaire dans :

$$\Delta x_t = \phi x_{t-1} + c + \sum_{i=1}^2 \gamma_i \Delta x_{t-i} + \mu_t \quad (4.163)$$

avec  $\mu_t$  *i.i.d.*  $(0, \sigma_\mu^2)$ . Si dans la procédure Eviews, on inclut uniquement une constante (*Intercept*) et l'on choisit un nombre de termes en différences retardés (*Lagged Difference*) égal à 2, on obtient les résultats suivants (figure 4.18).

Figure 4.18: Test de Racine Unitaire *ADF* sur CONSO : Modèle 2

ADF Test Statistic	1.151994	1% Critical Value*	-3.5031
		5% Critical Value	-2.8932
		10% Critical Value	-2.5834

\*MacKinnon critical values for rejection of hypothesis of a unit root.

Augmented Dickey-Fuller Test Equation  
 Dependent Variable: D(CONSO)  
 Method: Least Squares  
 Date: 09/24/01 Time: 10:41  
 Sample: 1978:4 2001:2  
 Included observations: 91

Variable	Coefficient	Std. Error	t-Statistic	Prob.
CONSO(-1)	0.007270	0.006311	1.151994	0.2525
D(CONSO(-1))	-0.152570	0.104192	-1.464310	0.1467
D(CONSO(-2))	0.246887	0.101993	2.420632	0.0176
C	-0.507829	0.942751	-0.538667	0.5915
R-squared	0.110958	Mean dependent var	0.652692	
Adjusted R-squared	0.080302	S.D. dependent var	0.994087	
S.E. of regression	0.953339	Akaike info criterion	2.785268	
Sum squared resid	79.07037	Schwarz criterion	2.895635	
Log likelihood	-122.7297	F-statistic	3.619394	
Durbin-Watson stat	2.099998	Prob(F-statistic)	0.016254	

La statistique de Student  $t_{\hat{\phi}}$  associée à la variable endogène retardée CONSO(-1) prend ici une valeur de 1.151. Pour tester l'hypothèse  $\phi = 0$ , on utilise alors les seuils tabulés par Dickey et Fuller pour le modèle 2 (*cf.* document joint, photocopié d'exercices) et pour une taille d'échantillon de 91 observations. Ces seuils sont reportés dans l'affichage de la procédure d'Eviews. Au seuil de 5%, le seuil critique est  $C_{(\alpha)} = -2.8932$  (-2.89 dans la table fournie en annexe pour 100 observations). Ainsi, dans ce cas pour un niveau de risque de 5%,  $t_{\hat{\phi}} > C_{(\alpha)}$ , on accepte *l'hypothèse nulle de racine unitaire* ( $\phi = 0$ ) dans le modèle 2.

Il faut à nouveau évaluer la validité de notre diagnostic en vérifiant que le modèle 2 à partir duquel nous avons fait le test de racine unitaire est bien le "bon" modèle. On teste pour cela la

nullité du coefficient de la constante conditionnellement à la présence d'une racine unitaire : test  $H_0^2$ ,  $(c, \phi) = (0, 0)$ . Dans le cas des tests de Dickey Fuller Augmentés, le test  $H_0^2$  ne correspond plus à l'hypothèse de nullité jointe de l'ensemble des coefficients de la régression. En effet, sous  $H_0^2$ , les coefficients  $\gamma_i$  ne sont pas nécessairement nuls. On ne peut donc plus utiliser la statistique de Fisher programmé dans Eviews pour le test de la nullité de l'ensemble des coefficients du modèle. On doit donc programmer la statistique de Fisher  $F_2$  associée à  $H_0^2$ .

```

!- CONSTRUCTION DES DIFFERENCES PREMIERES
SMPL 1978:2 2001:2
GENR DCONSO = CONSO-CONSO(-1)
!- ESTIMATION DU MODELE LIBRE
EQUATION MOD2.LS DCONSO C CONSO(-1) DCONSO(-1) DCONSO(-2)
SCALAR SCR2=@SSR
SCALAR NDL=@REGOBS-@NCOEF
!- ESTIMATION DU MODELE CONTRAINT
EQUATION MOD2.LS DCONSO DCONSO(-1) DCONSO(-2)
SCALAR SCR2C=@SSR
!- CONSTRUCTION DE LA STATISTIQUE F2
SCALAR F2=((SCR2C-SCR2)/2)/(SCR2/NDL)

```

La réalisation de la statistique de Fisher  $F_2$ , est égale à 8.47. On compare cette réalisation de  $F_2$  aux seuils critiques de la table de Dickey et Fuller (1981), tableau VI, page 1063, fournie en annexe (cf. polycopié d'exercices). Pour une taille d'échantillon de 100, et un risque de première espèce de 5%, la valeur critique est égale à 4.71. Pour un risque de 5%, la réalisation de  $F_3$  est supérieure au seuil critique, on *rejette donc l'hypothèse nulle de la nullité de la constante conditionnellement à la présence d'une racine unitaire*. Ceci signifie que le modèle 2 est le "bon" modèle, ce qui valide par la même notre diagnostic quant à la non stationnarité de la série de consommation.

Finalement, l'application des tests ADF nous indique que la série trimestrielle de consommation des ménages est engendrée par un processus non stationnaire  $I(1)$  de type  $AR(3)$ , puisque :

$$\begin{aligned}
 Conso_t = & \underset{(3.94)}{0.56} + Conso_{t-1} - \underset{(-1.26)}{0.13} (Conso_{t-1} - Conso_{t-2}) \\
 & + \underset{(2.56)}{0.26} (Conso_{t-2} - Conso_{t-3}) + \mu_t
 \end{aligned} \tag{4.164}$$

Par contre, la série de consommation différenciée (qui correspond au taux de croissance si la série en logarithme), est générée par un processus stationnaire  $I(0)$  de type  $AR(2)$  avec constante :

$$\Delta Conso_t = \underset{(3.94)}{0.56} - \underset{(-1.26)}{0.13} \Delta Conso_{t-1} + \underset{(2.56)}{0.26} \Delta Conso_{t-2} + \mu_t \tag{4.165}$$

où  $\mu_t$  est un processus bruit blanc. On vérifie en effet que le choix du lag  $p = 2$ , nous a permis de blanchir totalement les résidus  $\mu_t$ , comme le confirme l'examen du corrélogramme (non reporté).

## 5 Les processus ARIMA

Il ne nous reste plus qu'à introduire à présent une sous-classe de processus ARMA, la classe des processus ARIMA c'est les processus ARMA intégré ou DS. Nous distinguerons les processus ARIMA non saisonniers et les processus saisonniers SARIMA

**Definition 29** *Le processus stationnaire  $x_t$  satisfait une représentation ARIMA (Integrated AutoRegressive Moving Average) d'ordre  $p$  et  $q$ , intégré d'ordre  $d$ , notée ARIMA( $p, d, q$ ), si :*

$$\Phi(L)(1-L)^d x_t = c + \Theta(L)\varepsilon_t \quad (5.166)$$

avec  $c \in \mathbb{R}$ ,  $\Theta(L) = \sum_{j=0}^q \theta_j L^j$ ,  $\Phi(L) = \sum_{j=0}^p \phi_j L^j$  où  $\forall j < q \theta_j \in \mathbb{R}^2, \forall j < p \phi_j \in \mathbb{R}^2$ ,  $\theta_0 = \phi_0 = 1$  et  $(\phi_p, \theta_q) \in \mathbb{R}^{2*}$ , avec  $\{\varepsilon_t\}$  i.i.d.  $(0, \sigma_\varepsilon^2)$ . Les polynômes  $\Theta(L)$  et  $\Phi(L)$  ont toutes racines situées à l'extérieur du cercle unité.

**Definition 30** *Le processus stationnaire  $x_t$  satisfait une représentation SARIMA saisonnier (Seasonal Integrated AutoRegressive Moving Average) d'ordre  $p$  et  $q$ , intégré d'ordre  $d$ , de période  $s$  et  $s'$ , notée SARIMA $_{s,s'}(p, d, q)$ , si :*

$$\Phi(L^s)(1-L^s)^d x_t = c + \Theta(L^{s'})\varepsilon_t \quad (5.167)$$

avec  $c \in \mathbb{R}$ ,  $\Theta(L) = \sum_{j=0}^q \theta_j L^j$ ,  $\Phi(L) = \sum_{j=0}^p \phi_j L^j$  où  $\forall j < q \theta_j \in \mathbb{R}^2, \forall j < p \phi_j \in \mathbb{R}^2$ ,  $\theta_0 = \phi_0 = 1$  et  $(\phi_p, \theta_q) \in \mathbb{R}^{2*}$ , avec  $\{\varepsilon_t\}$  i.i.d.  $(0, \sigma_\varepsilon^2)$ . Les polynômes  $\Theta(L)$  et  $\Phi(L)$  ont toutes racines situées à l'extérieur du cercle unité.

## A Annexes

### A.1 Simulations de régressions fallacieuses

Le programme des simulations est le suivant :

```

'TSTATISTIQUES : SPURIOUS REGRESSIONS
VECTOR (1000) TSTAT5000
VECTOR (1000) TSTAT1000
VECTOR (1000) TSTAT500
VECTOR (1000) TSTAT100
FOR !I=1 TO 1000
  'INITIALISATION DU PROCESSUS
  SMPL 1 5000
  GENR X=0
  GENR Y=0
  ' CONSTRUCTION DES MARCHES ALÉATOIRES
  SMPL 2 5000
  GENR X=X(-1)+NRND
  GENR Y=Y(-1)+NRND
  ' REGRESSION 5000 POINTS
  SMPL 1 5000
  EQUATION EQ1.LS Y C X
  TSTAT5000(!I)=EQ1.@TSTATS(1)
  ' REGRESSION 1000 POINTS
  SMPL 1 1000
  EQUATION EQ1.LS Y C X
  TSTAT1000(!I)=EQ1.@TSTATS(1)
  ' REGRESSION 500 POINTS
  SMPL 1 500
  EQUATION EQ1.LS Y C X
  TSTAT500(!I)=EQ1.@TSTATS(1)
  ' REGRESSION 100 POINTS
  SMPL 1 100
  EQUATION EQ1.LS Y C X
  TSTAT100(!I)=EQ1.@TSTATS(1)
NEXT
SMPL 1 100
MTOS(TSTAT5000,TSTAT5000_s)

```

MTOS(TSTAT1000,TSTAT1000\_s)

MTOS(TSTAT500,TSTAT500\_s)

MTOS(TSTAT100,TSTAT100\_s)