

Chapitre 2:
La non stationnarité - Tests de détection
Quelques notes de cours (non exhaustives)

1 Définition de la non stationnarité

La plupart des séries économiques sont non stationnaires, c'est-à-dire que le processus qui les décrit ne vérifie pas au moins une des conditions de la définition d'un processus stationnaire du second ordre, donnée par :

- $E(Y_t) = m$ indépendant du temps
- $V(Y_t) = \gamma(0) < \infty$, $\gamma(0)$ indépendant du temps
- $Cov(Y_t, Y_{t-h}) = \gamma(h)$ ne dépend pas de t

Ceci nous conduit à définir deux types de non stationnarité selon que c'est plutôt la condition portant sur le moment d'ordre 1 qui n'est pas vérifiée (non stationnarité déterministe) ou les conditions portant sur les moments du second ordre qui ne sont pas vérifiées (non stationnarité stochastique).

1.1 Non stationnarité déterministe

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

$$Y_t = f(t) + Z_t$$

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

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

$$Y_t - f(t) = Z_t \text{ stationnaire}$$

Le processus Z_t peut être modélisé par un processus $ARMA(p, q)$ stationnaire (le bruit blanc étant un cas particulier), admettant une représentation canonique $\Phi(L)Z_t = \Theta(L)\varepsilon_t$.

$f(t)$ est une fonction déterministe, par exemple $f(t) = a + bt$ (cas le plus couramment retenu), mais on pourrait aussi considérer, entre autres, un tendance quadratique $f(t) = a + bt + ct^2$ ou une tendance segmentée $f(t) = a_1 + b_1t + a_2\mathbb{I}_{t < t_0} + b_2t\mathbb{I}_{t < t_0}$, avec $\mathbb{I}_{t < t_0} = 1$ si $t < t_0$ et 0 sinon.

Une première conséquence économique (d'avoir un processus TS) est qu'un choc imprévu (ε_t) n'a pas d'effet persistant sur le processus puisqu'il ne peut pas modifier sa partie tendancielle (sa croissance), qui est ici exogène. Il n'aura donc d'effet que sur la partie cyclique, supposée être stationnaire, donc son effet sera forcément temporaire. Une deuxième conséquence économique est que la décomposition tendance-cycle est naturelle dans ce cas : la tendance est donnée par $f(t)$ et le cycle par les écarts de la série à sa tendance, soit Z_t . Les deux composantes ne sont pas corrélées.

1.2 Non stationnarité stochastique

On dit que le processus Y_t est caractérisé par une non stationnarité stochastique, ou encore que le processus Y_t est DS (*Difference stationary*) si le processus différencié une fois $(1 - L)Y_t$ est stationnaire. On parle aussi de processus intégré d'ordre 1, on note $Y_t \sim I(1)$:

$$(1 - L)Y_t = Z_t \text{ stationnaire} \implies Y_t = Y_{t-1} + Z_t$$

De manière générale, on dit que le processus Y_t est un processus intégré d'ordre d , avec d le degré d'intégration, si le processus différencié d fois $(1 - L)^d Y_t$ est stationnaire. On note $Y_t \sim I(d)$:

$$(1 - L)^d Y_t = Z_t$$

Si Z_t suit un modèle $ARMA(p, q)$ stationnaire, $\Phi_p(L)Z_t = \Theta(L)\varepsilon_t$, avec $\Phi_p(L)$ un polynôme de degré p dont les racines sont toutes strictement supérieures à 1 en module, $\Theta(L)$ un polynôme de degré q dont les racines sont toutes strictement supérieures à 1 en module, et ε_t un bruit blanc $(0, \sigma^2)$:

$$\Phi_p(L)(1 - L)^d Y_t = \Theta(L)\varepsilon_t \iff \Phi_{p+d}(L)Y_t = \Theta(L)\varepsilon_t$$

avec $\Phi_{p+d}(L)$ un polynôme de degré $p + d$ avec d racines égales à 1 et les p autres strictement supérieures à 1 en module.

Ainsi, dire qu'un processus est $I(d)$ signifie que le polynôme en L définissant sa partie autorégressive a d racines unitaires.

Les exemples les plus connus de processus $I(1)$ sont, d'une part, la marche aléatoire pure :

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

et, d'autre part, la marche aléatoire avec dérive :

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

Il s'agit d'un AR(1) avec racine unitaire. Ainsi, on parle aussi de présence de racine(s) unitaire(s) pour parler des processus non stationnaires de type stochastique.

Une première conséquence importante (d'avoir un processus DS) est qu'un choc imprévu (ε_t) à une date donnée influence la tendance et le futur du processus. Le processus est caractérisé par de la persistance des chocs ou de l'hystérèse. Autrement dit, un choc temporaire à une date donnée a un effet permanent sur le niveau du processus puisque le processus ne rejoindra jamais sa valeur initiale suite à ce choc. Une deuxième conséquence est que la décomposition tendance-cycle n'est plus explicite dans cette formulation. On peut l'obtenir par des méthodes de décomposition adaptées (décomposition de Beveridge et Neslon par exemple).

2 Détection de la non stationnarité et de sa nature

La représentation graphique de la série peut (dans certains cas) nous indiquer que la série n'est pas stationnaire, mais elle ne nous permet pas de discriminer entre les deux types de non

stationnarité (la présence de non stationnarité de type stochastique peut produire une série à l'allure croissante, de la même manière que la non stationnarité de type déterministe).

Quand la série est non stationnaire (que ce soit DS ou TS), le coefficient d'autocorrélation d'ordre 1 (empirique) est très élevé (proche de 1), l'autocorrélogramme simple décroît alors lentement. Ainsi, l'autocorrélogramme simple d'un processus non stationnaire est celui d'un AR(1) avec ϕ proche de 1, que le processus soit TS ou DS. Rappelons que l'autocorrélogramme d'un processus stationnaire présente une décroissance "rapide", puisque les autocorrélations simples $\rho(h) \rightarrow 0$ quand $h \rightarrow \infty$.

De même, on observe un pic à l'origine dans la représentation de la densité spectrale pour un processus non stationnaire.

Ainsi, la représentation graphique, l'autocorrélogramme simple et la densité spectrale peuvent nous indiquer la présence de non stationnarité, mais ne peuvent pas nous renseigner sur son type (DS ou TS).

Les deux types de non stationnarité, DS et TS, n'ayant pas les mêmes conséquences statistiques sur la nature du processus, et nécessitant des méthodes différentes pour rendre le processus stationnaire, on peut mettre en œuvre une méthode, que l'on peut qualifier de méthode heuristique, pour se donner une idée de la nature de la non stationnarité. Il s'agit de stationnariser la série, d'une part par différenciation, d'autre part par retrait d'une composante déterministe, et d'observer, à l'aide de la représentation graphique et de l'autocorrélogramme, les résultats obtenus. Cette méthode est basée sur les conséquences de l'application d'une méthode de stationnarisation à tort.

2.1 Méthode heuristique

En différenciant une série TS, on rend l'allure de la série stationnaire, mais on crée une racine unitaire dans la partie MA (le processus différencié obtenu n'ayant alors plus une représentation canonique). En effet, supposons que le processus Y_t soit de type TS défini par :

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

Si on applique la méthode de stationnarisation d'un processus I(1), soit le filtre $1 - L$:

$$(1 - L)Y_t = (1 - L)(a + bt + \varepsilon_t) = b + \varepsilon_t - \varepsilon_{t-1}$$

ceci conduit, certes à stationnariser la série puisque $E((1 - L)Y_t) = b$, $V((1 - L)Y_t) = 2\sigma^2$ et $cov((1 - L)Y_t, (1 - L)Y_{t-h}) = -\sigma^2$ si $h = \pm 1$ et 0 sinon, mais cela introduit une racine unitaire dans la partie MA du processus différencié.

En retirant une composante déterministe à une série DS, on rend (sans doute) l'allure de la série stationnaire, mais on ne retire pas la racine unitaire dans la partie AR, ce qui conduit à conserver une forte autocorrélation (visible dans l'autocorrélogramme simple). En effet, supposons que le processus soit DS, de type I(1) :

$$Y_t = Y_{t-1} + \varepsilon_t \Leftrightarrow Y_t = Y_0 + \sum_{j=0}^{t-1} \varepsilon_{t-j}$$

Si on applique la méthode de stationnarisation d'un processus TS, on risque d'obtenir quelque chose du genre :

$$Y_t - f(t) \simeq \sum_{j=0}^{t-1} \varepsilon_{t-j}$$

la série obtenue n'est toujours pas stationnaire puisque le problème de non constance de la variance n'est pas corrigé.

Ainsi, l'extraction d'une tendance déterministe à un processus DS conduit à créer artificiellement une forte autocorrélation des résidus aux premiers retards.

2.2 Tests de racine unitaire: les tests de Dickey-Fuller

Un test de non stationnarité largement utilisé et répandu est le test de racine unitaire proposé par Dickey et Fuller en 1979. L'hypothèse nulle du test est la présence de racine unitaire, soit la non stationnarité de type stochastique.

Le test consiste à tester :

$$H_0 : \phi = 1$$

contre

$$H_1 : \phi < 1$$

dans le modèle

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

avec ε_t bruit blanc ($0, \sigma^2$). L'hypothèse nulle correspond au cas de marche aléatoire pure (processus DS, I(1)) et l'hypothèse alternative correspond au cas d'un modèle AR(1) stationnaire.

Pour mener ce test, on calcule la statistique de Student, mais attention, cette statistique ne suit plus sous l'hypothèse nulle une loi de Student, puisque, sous l'hypothèse nulle, le processus est non stationnaire de type DS et les propriétés asymptotiques ne sont plus standards. Ainsi, la différence avec un test standard repose sur les valeurs critiques à utiliser pour conclure sur le test. On ne peut plus utiliser 1.64 comme valeur critique pour un test unilatéral à 5%. Il faut utiliser les valeurs critiques, qui ont été retabulées par Dickey et Fuller.

De plus, ce test ne répond pas à nos attentes de détection du type de non stationnarité dans les variables économiques, d'une part parce que l'hypothèse de processus TS n'est pas présente et d'autre part parce que les séries économiques sont caractérisées par de l'autocorrélation, qui conduira la plupart du temps à rejeter l'hypothèse de bruit blanc pour ε_t dans le test ci-dessus. Pour prendre en compte, d'une part la présence d'autocorrélation dans les séries économiques, et, d'autre part, l'hypothèse de tendance déterministe, on mène les tests de racine unitaire dans les trois régressions suivantes.

$$\Delta Y_t = \rho Y_{t-1} + \alpha + \beta t + \sum_{j=1}^p \phi_j \Delta Y_{t-j} + \varepsilon_t \quad (1)$$

$$\Delta Y_t = \rho Y_{t-1} + \alpha + \sum_{j=1}^p \phi_j \Delta Y_{t-j} + \varepsilon_t \quad (2)$$

$$\Delta Y_t = \rho Y_{t-1} + \sum_{j=1}^p \phi_j \Delta Y_{t-j} + \varepsilon_t \quad (3)$$

avec p le nombre de retards à ajouter dans la régression afin de prendre en compte l'autocorrélation et donc de “blanchir” les résidus. On parle de correction paramétrique de l'autocorrélation et on appelle les tests de Dickey-Fuller, les tests de Dickey-Fuller augmenté (ADF).

Le test de racine unitaire consiste alors à tester :

$$H_0 : \rho = 0$$

contre

$$H_1 : \rho < 0$$

dans les modèles (1), (2) et (3).

- Dans le modèle (1) :

ΔY_t est $I(0) + T$ (il a une tendance déterministe et l'écart à cette tendance déterministe suit un modèle $AR(p)$ stationnaire) sous H_0 , c'est-à-dire que Y_t est $I(1) + T^2$.

Sous H_1 , Y_t a une tendance déterministe et l'écart à cette tendance déterministe suit un modèle AR stationnaire, on note $I(0) + T$, soit un processus TS.

- Dans le modèle (2) :

ΔY_t est $I(0) + C$ (il suit un modèle $AR(p)$ stationnaire non centré) sous H_0 , c'est-à-dire que Y_t est $I(1) + T$.

Sous H_1 , Y_t suit un modèle AR stationnaire non centré, on note $I(0) + C$.

- Dans le modèle (3) :

ΔY_t est $I(0)$ (il suit un modèle $AR(p)$ stationnaire centré) sous H_0 , c'est-à-dire que Y_t est $I(1)$.

Sous H_1 , Y_t suit un modèle AR stationnaire, on note $Y_t \sim I(0)$.

Les valeurs critiques du test de racine unitaire dépendent de la présence ou non d'une constante ou d'une tendance. Ainsi, il faut comparer la statistique de student à la valeur critique pertinente. Soit, au seuil de 5%, on doit considérer comme valeur critique, non pas -1.64 comme dans le cas standard, mais -3.45 dans le modèle (3), -2.89 dans le modèle (2) et -1.95 dans le modèle (1). Sous SAS, les tables sont intégrées dans le logiciel, et SAS donne directement les p-values calculées selon les bonnes valeurs critiques!

De plus, puisque les valeurs critiques dépendent de la présence ou non d'une constante ou d'une tendance, cela implique que le test de racine unitaire doit être mené dans le “bon” modèle.

Ainsi, une possibilité de mettre en œuvre les tests de racine unitaire est de procéder de manière emboîtée, selon la stratégie suivante : on teste la racine unitaire dans le modèle le plus général, puis on teste si le modèle utilisé pour mener le test était pertinent. Si tel n'est pas le cas, on doit mener à nouveau le test de racine unitaire dans le modèle contraint, etc.

Une stratégie de test de racine unitaire :

- (i) On choisit le nombre de retards p à introduire dans la régression : on peut, pour cela, choisir l'ordre p de l' $AR(p)$ pour la variable $(1-L)Y_t$, sur la base des autocorrélations partielles de $(1-L)Y_t$, et sur la base de la significativité du dernier retard de l'AR introduit dans la régression, tout en vérifiant que le résidu est bien un bruit blanc.
- (ii) On teste la racine unitaire $H_0 : \rho = 0$ dans le modèle le plus général (3) :

$$\Delta Y_t = \rho Y_{t-1} + \alpha + \beta t + \sum_{j=1}^p \phi_j \Delta Y_{t-j} + \varepsilon_t$$

- Si on accepte H_0 (la racine unitaire), alors on va tester :

$$H_{01} : \rho = 0 \text{ et } \beta = 0$$

par un test de Fisher (dont les valeurs critiques ne sont pas standards) afin de s'assurer que le test de racine unitaire a été mené dans le bon modèle.

- * Si on accepte H_{01} , alors on passe à l'étape (iii)
 - * Si on rejette H_{01} , on conclut que le processus est $I(1) + T^2$ (tout en sachant que cette conclusion est peu crédible économiquement, et qu'elle cache peut être le cas d'une tendance déterministe plus complexe que linéaire)
- Si on rejette H_0 (rejet de la racine unitaire), alors le processus est stationnaire, mais on doit aller tester la pertinence d'avoir testé la racine unitaire dans un modèle avec tendance en testant la significativité de la tendance (par un student normal) dans le modèle suivant :

$$\Delta Y_t = \rho Y_{t-1} + \alpha + \beta t + \sum_{j=1}^p \phi_j \Delta Y_{t-j} + \varepsilon_t$$

- * Si on accepte $H_0 : \beta = 0$, alors il est recommandé de passer à l'étape (iii)
 - * Si on rejette $H_0 : \beta = 0$, on conclut que le processus est $I(0) + T$
- (iii) On teste la racine unitaire $H_0 : \rho = 0$ dans le modèle (2) :

$$\Delta Y_t = \rho Y_{t-1} + \alpha + \sum_{j=1}^p \phi_j \Delta Y_{t-j} + \varepsilon_t$$

- Si on accepte H_0 (la racine unitaire), alors on va tester :

$$H_{02} : \rho = 0 \text{ et } \alpha = 0$$

par un test de Fisher (dont les valeurs critiques ne sont pas standards) afin de s'assurer que le test de racine unitaire a été mené dans le bon modèle.

- * Si on accepte H_{02} , alors on passe à l'étape (iv)
- * Si on rejette H_{02} , on conclut que le processus est $I(1) + T$ (marche aléatoire avec dérive).

- Si on rejette H_0 (rejet de la racine unitaire), alors le processus est stationnaire, mais on doit aller tester la pertinence d'avoir testé la racine unitaire dans un modèle avec constante en testant la significativité de la constante (par un student normal) dans le modèle suivant :

$$\Delta Y_t = \rho Y_{t-1} + \alpha + \sum_{j=1}^p \phi_j \Delta Y_{t-j} + \varepsilon_t$$

- * Si on accepte H_0 : $\alpha = 0$, alors il est recommandé de passer à l'étape (iv)
- * Si on rejette H_0 : $\alpha = 0$, on conclut que le processus est $I(0) + C$

(iv) On teste la racine unitaire H_0 : $\rho = 0$ dans le modèle (3) :

$$\Delta Y_t = \rho Y_{t-1} + \sum_{j=1}^p \phi_j \Delta Y_{t-j} + \varepsilon_t$$

- Si on accepte H_0 (la racine unitaire), alors le processus est $I(1)$
- Si on rejette H_0 (rejet de la racine unitaire), alors le processus est stationnaire centré $I(0)$

Table 1: Valeurs critiques des tests de racine unitaire de Dickey et Fuller

Modèles	10%			5%			1%		
	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
T=50	-3.18	-2.6	-1.61	-3.5	-2.93	-1.95	-4.15	-3.58	-2.62
T=100	-3.15	-2.58	-1.61	-3.45	-2.89	-1.95	-4.04	-3.51	-2.60
T=250	-3.13	-2.57	-1.62	-3.43	-2.88	-1.95	-3.99	-3.46	-2.58

2.3 Tests de racine unitaire: les tests de Phillips-Perron

Phillips-Perron (1988) proposent une méthode non paramétrique pour corriger de la présence d'autocorrélation, sans avoir à ajouter des endogènes retardées comme dans la méthode de DF augmentée (méthode plus robuste en cas d'erreurs MA notamment).

La procédure de test consiste à tester l'hypothèse de racine unitaire H_0 : $\rho = 0$ dans les modèles suivants :

$$\Delta Y_t = \rho Y_{t-1} + \alpha + \beta t + \varepsilon_t$$

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

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

La statistique de test de Phillips-Perron (PP) est une statistique de student corrigée de la présence d'autocorrélation par la prise en compte d'une estimation de la variance de long terme de ε_t (calculée par la densité spectrale de ε_t à la fréquence zéro), robuste à la présence d'autocorrélation et d'hétéroscédasticité.

Cette estimation de la variance de long terme de ε_t est donnée par :

$$w^2 = \gamma_\varepsilon(0) + 2 \sum_{j=1}^q \gamma_\varepsilon(j)$$

où $\gamma_\varepsilon(j)$, $j = 0, 1, \dots, q$ est le coefficient d'autocovariance d'ordre j de ε_t et q est un paramètre de troncature (pour ne pas calculer les coefficients d'autocovariances jusqu'à un ordre infini!), choisi par exemple de la manière suivante :

$$q = \text{floor} \left[4 \left(\frac{T}{100} \right)^{2/9} \right] \quad (= 4 \text{ si } T = 100; 5 \text{ si } T = 500)$$

On préfère en général utiliser la correction de Newey-West :

$$w^2 = \gamma_\varepsilon(0) + 2 \sum_{j=1}^q \left(1 - \frac{j}{q+1} \right) \gamma_\varepsilon(j)$$

La distribution asymptotique de la statistique de student corrigée de w est la même que celle de DF.