

Variables non stationnaires - les séries avec trend

Gilbert Colletaz

29 décembre 2020

Résumé

Nous allons traiter ici de variables non stationnaires, ce qui nous conduira notamment à considérer des variables avec tendance. On devra alors distinguer entre tendance déterministe et tendance stochastique compte-tenu des propriétés différentes qu'elles imposent sur les variables économiques qui les contiennent et aussi en raison des différences entre les modèles statistiques adaptés à chacune d'elles. Cette partie s'achèvera avec l'exposé des principaux tests de présence d'une racine unitaire.

Table des matières

1	La prise en compte d'une tendance	1
1.1	introduction aux séries à trend déterministe	2
1.2	l'estimation d'un trend déterministe linéaire	2
1.2.1	tendance linéaire augmentée d'un bruit blanc	3
1.2.2	processus AR(1) avec trend linéaire déterministe	4
1.3	prévisions dans les processus à trend déterministe	5
1.3.1	les chocs cycliques et aléatoires n'ont que des effets transitoires	5
1.3.2	la largeur des IC afférents aux prévisions de la non stationnaire est finie	5
2	Les processus à trend stochastique	6
2.1	processus à racine unitaire et processus intégrés	6
2.2	les processus ARIMA	6
2.2.1	processus à racine unitaire, processus intégrés, processus ARIMA	7
2.3	comparaison trend déterministe - trend stochastique : AR(1)-TS versus AR(1)-DS	7
2.4	mesure de l'impact d'un choc d'innovation sur une série I(1)	8
3	Les tests de racine unitaire	10
3.1	les tests DF de présence d'une racine unitaire dans le polynôme caractéristique d'un AR(1)	11
3.1.1	Les statistiques τ , τ_μ et τ_t de Dickey-Fuller	11
3.1.2	La recherche des valeurs critiques des tests DF	13
3.2	le test ADF : racine unitaire dans le polynôme AR d'un AR(p) ou d'un ARMA(p,q), $p > 0$	14
3.2.1	test de présence d'une racine unitaire	14
3.2.2	à la recherche d'une autre racine unitaire	16
3.2.3	statistiques similaires	16
3.3	le test de Phillips-Perron	17

1 La prise en compte d'une tendance

Jusqu'ici nous avons supposé que nous traitons des variables stationnaires et le plus souvent stationnaires au sens large, *i.e.* telles que, pour tout t ,

$$E[x_t] = \mu_X, \text{ indépendant de } t$$
$$Cov(x_t, x_{t-h}) = \gamma_h, \text{ indépendant de } t$$

Or on sait que l'on peut repérer sur les valeurs de nombreuses variables économiques des tendances haussières ou baissières qui subsistent trop longtemps pour que l'on puisse raisonnablement penser que le niveau moyen est identique en

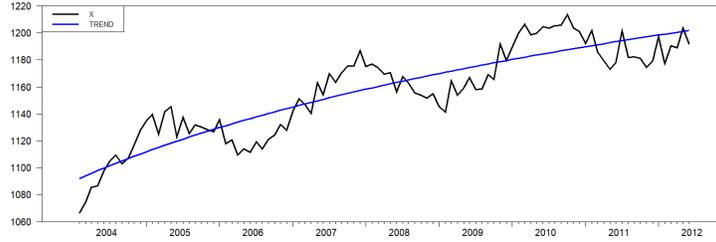


FIGURE 1 – Illustration d’une série avec tendance

début et en fin de période. Dans ces conditions, sur ces variables, il faut au moins corriger la première des deux équations ci-dessus pour poser maintenant quelque chose comme : $E[x_t] = \mu_t$.

Vous devez comprendre que si les moments dépendent du temps, alors tous les outils d’identification et d’estimation précédemment étudiés sont susceptibles de ne plus servir à rien. En effet, dans ce cas la structure d’autocovariance de la série, et donc ses autocorrélations et ses autocorrélations partielles, vont se transformer avec t . En pratique, on n’aura plus par exemple un échantillon de taille T pour estimer la moyenne de x , mais un échantillon de taille 1 pour estimer une espérance, μ_t , qui varie à chaque nouvelle observation.

1.1 introduction aux séries à trend déterministe

L’approche la plus simple traduisant l’idée d’une espérance qui se modifie avec t est d’admettre que les valeurs des observations de x sont en partie déterminées par celles d’une autre variable, le trend, dont le niveau est dépendant du temps. La figure 1 illustre le raisonnement qui précède : le trend y est responsable d’un mouvement haussier sur les années 2004-2012. Par ailleurs, on peut également observer une alternance de sous-périodes de 2-3 ans pendant lesquelles la variable x est soit au-dessus, soit en-dessous du trend. Ces sous-périodes témoignent également de la présence d’un cycle stationnaire dans la mesure où il ne provoque pas de divergence durable par rapport au niveau moyen donné par le trend. Enfin, on voit que x n’est pas une série parfaitement lisse. Cette irrégularité peut être mise sur le compte d’une composante aléatoire stationnaire indépendante du trend et du cycle. On a ainsi une décomposition de la forme ¹ :

$$x_t = T_t + c_t + u_t \tag{1}$$

où $T_t = f(t)$ est la valeur d’un trend, c_t celle du cycle et u_t l’aléa résiduel, les deux dernières variables étant stationnaires. Ces processus, où la variable d’intérêt, x_t , peut être vue comme une variable stationnaire oscillant autour d’un trend non stationnaire sont qualifiés de *trend stationary*.

Remarquez aussi que l’on peut poser sans perte de généralité que les moyennes des composantes cyclique et aléatoire peuvent être considérées comme nulles. Pour le montrer, supposons que cela ne soit pas vrai : $x_t = \tilde{T}_t + \tilde{c}_t + \tilde{u}_t$ avec $E[\tilde{c}_t] = \mu_{\tilde{c}} \neq 0$ et $E[\tilde{u}_t] = \mu_{\tilde{u}} \neq 0$ où \tilde{T}_t est un trend déterministe. Alors, en posant $T_t = \tilde{T}_t + \mu_{\tilde{c}} + \mu_{\tilde{u}}$, on retrouve bien $x_t = T_t + c_t + u_t$ avec $E[c_t] = E[u_t] = 0$.

On note encore qu’avec cette décomposition, si on connaît le trend alors les nouvelles difficultés disparaissent en partie puisqu’il suffit de créer par simple différence une variable transformée $\tilde{x}_t = x_t - T_t = c_t + u_t$ pour récupérer une variable stationnaire sur laquelle les outils d’identification et d’estimation peuvent valablement opérer. Il serait alors possible de rechercher un filtre optimal sur \tilde{x}_t pour en construire des prévisions ${}_t\tilde{x}_{t+h}$.

1.2 l’estimation d’un trend déterministe linéaire

On traite dans cette partie de l’estimation des processus incorporant un trend déterministe. Notez que ce trend n’est pas obligatoirement linéaire, par exemple :

- $T_t = a + bt$,

1. En économie le graphique 1 est quelquefois utilisé pour illustrer le découpage des enseignements entre une théorie de la croissance qui viserait à expliquer le trend de long terme, par exemple via la productivité des facteurs de production et leurs volumes, et la théorie keynésienne qui serait plus attachée à comprendre et agir sur les aspects plus conjoncturels de l’activité économiques.

- $T_t = at^2 + bt + c$,
- $T_t = \log(t)$
- $T_t = \sqrt{T}$

sont des fonctions qui, si les paramètres a,b,c sont connus permettent toutes de faire des prévisions sans erreur sur les valeurs futures de la tendance. Mais la première est certainement la plus simple à manipuler et c'est donc elle que nous allons utiliser.

1.2.1 tendance linéaire augmentée d'un bruit blanc

C'est la configuration la plus simple, il n'y a pas de cycle et la partie aléatoire est un bruit blanc :

$$x_t = a + bt + u_t, \quad b \neq 0 \quad (2)$$

L'étude des propriétés des estimateurs OLS de cette équation ² n'est pas triviale, même si la mise en évidence des difficultés est simple. En effet, si on ajuste l'équation précédente par OLS, il vient :

$$\begin{pmatrix} \hat{a} \\ \hat{b} \end{pmatrix} = \begin{pmatrix} a \\ b \end{pmatrix} + \begin{pmatrix} T & \sum_{t=1}^T t \\ \sum_{t=1}^T t & \sum_{t=1}^T t^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum_{t=1}^T ut \\ \sum_{t=1}^T tu_t \end{pmatrix}$$

Il faut chercher une normalisation pour que

$$\begin{pmatrix} T & \sum_{t=1}^T t \\ \sum_{t=1}^T t & \sum_{t=1}^T t^2 \end{pmatrix}$$

converge vers une matrice de constantes lorsque $T \rightarrow \infty$. Or :

$$\lim_{t \rightarrow \infty} \sum_{t=1}^T t = \frac{T(T+1)}{2} \quad \text{et} \quad \lim_{t \rightarrow \infty} \sum_{t=1}^T t^2 = \frac{T(T+1)(2T+1)}{6}$$

d'où :

$$\begin{aligned} \lim_{t \rightarrow \infty} \frac{1}{T^2} \sum_{t=1}^T t &= \frac{1}{2}, \quad \text{et} \quad \lim_{t \rightarrow \infty} \frac{1}{T^2} \sum_{t=1}^T t^2 = \infty \\ \lim_{t \rightarrow \infty} \frac{1}{T^3} \sum_{t=1}^T t^2 &= 0, \quad \text{et} \quad \lim_{t \rightarrow \infty} \frac{1}{T^3} \sum_{t=1}^T t^3 = \frac{1}{3} \end{aligned}$$

Ainsi, si on normalise par $1/T^2$ on obtient une matrice qui diverge et si on normalise par $1/T^3$ on obtient une matrice qui converge vers une matrice de constantes non inversible : Il faut donc appliquer une normalisation différente ce qui implique alors des vitesses de convergence différentes pour \hat{a} et \hat{b} . Finalement, on peut montrer que :

$$\begin{pmatrix} T^{1/2}(\hat{a} - a) \\ T^{3/2}(\hat{b} - b) \end{pmatrix} \rightarrow \mathcal{N} \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \sigma_u^2 \begin{pmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{3} \end{pmatrix}^{-1} \right) = \mathcal{N} \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \sigma_u^2 \begin{pmatrix} 4 & -6 \\ -6 & 12 \end{pmatrix} \right) \quad (3)$$

On remarque que l'estimateur de la constante de la régression possède la vitesse usuelle en \sqrt{T} alors que l'estimateur de la pente converge plus vite : \hat{b} est qualifié de *super-convergent*.

Ainsi toujours asymptotiquement,

$$T^{1/2}(\hat{a} - a) \rightarrow \mathcal{N}(0, 4\sigma_u^2) \Rightarrow (\hat{a} - a) \rightarrow \mathcal{N}(0, 4\sigma_u^2/T) \quad (4)$$

et

$$T^{3/2}(\hat{b} - b) \rightarrow \mathcal{N}(0, 12\sigma_u^2) \Rightarrow (\hat{b} - b) \rightarrow \mathcal{N}(0, 12\sigma_u^2/T^3) \quad (5)$$

2. Ici le trend linéaire déterministe explique le niveau de la variable x . Une autre formulation également rencontrée pose que le trend explique le logarithme de x , soit $y_t = \log(x_t)$ et $y_t = a + bt + u_t$. Cette dernière version implique le taux de croissance moyen de x est constant puisque $dy_t = \frac{dx_t}{x_t} = b dt$, et que x obéit à une trend exponentiel car $x_t = \exp(a + bt + u_t)$. Cette modélisation est par exemple souvent appliquée sur des séries macro-économiques comme le PIB, la consommation nationale, les importations, etc...

Dans ces conditions, avec l'estimateur usuel $s_u^2 = \sum_{t=1}^T \hat{u}_t^2 / (T-2)$ et comme $s_u^2 \rightarrow \sigma_u^2$, un test de student sur la constante de la régression vérifiera :

$$t_a = \frac{\hat{a} - a}{s_{\hat{a}}} = \frac{\sqrt{T}(\hat{a} - a)}{\sqrt{T}s_{\hat{a}}} = \frac{\sqrt{T}(\hat{a} - a)}{\sqrt{T} \left[2\sigma_u / \sqrt{T} \right]} = \frac{\sqrt{T}(\hat{a} - a)}{2\sigma_u} \rightarrow \mathcal{N}(0, 1) \text{ d'après (4)}$$

et que pour un test de student sur la pente,

$$t_b = \frac{\hat{b} - b}{s_{\hat{b}}} = \frac{T^{3/2}(\hat{b} - b)}{T^{3/2}s_{\hat{b}}} = \frac{T^{3/2}(\hat{b} - b)}{T^{3/2} \left[\sqrt{12}\sigma_u / T^{3/2} \right]} = \frac{T^{3/2}(\hat{b} - b)}{\sqrt{12}\sigma_u} \rightarrow \mathcal{N}(0, 1) \text{ d'après (5)}$$

Dans les deux cas on retrouve asymptotiquement des gaussiennes centrées réduites sous H_0 : on peut donc, malgré la non stationnarité et des vitesses de convergence différentes, utiliser les tests classiques d'hypothèse dans ce modèle ³.

1.2.2 processus AR(1) avec trend linéaire déterministe

C'est une extension immédiate du modèle précédent, et il sera utile lorsque nous voudrions comparer les processus à trend déterministe et les processus à trend stochastique. Soit pour $t = 0, 1, 2, \dots$:

$$x_t = \phi_0 + \phi_1 x_{t-1} + bt + u_t, \quad |\phi_1| < 1, \quad b \neq 0, \quad u_t \text{ bruit blanc} \quad (6)$$

C'est évidemment un processus non stationnaire. Par substitutions répétées des x retardés vous pouvez montrer que :

$$x_t = \phi_0 \sum_{i=0}^{t-1} \phi_1^i - b \sum_{i=1}^{t-1} i\phi_1^i + b \left(\sum_{i=1}^{t-1} \phi_1^i \right) \times t + \phi_1^t x_0 + \sum_{i=0}^{t-1} \phi_1^i u_{t-i} \quad (7)$$

et on peut alors définir l'écart de x_t à son trend, soit y_t comme :

$$y_t = x_t - b \left(\sum_{i=1}^{t-1} \phi_1^i \right) \times t, \quad (8)$$

Il vient encore

$$\lim_{t \rightarrow \infty} \left(\sum_{i=1}^{t-1} \phi_1^i \right) = \frac{\phi_1}{(1 - \phi_1)} \text{ et donc } \lim_{t \rightarrow \infty} y_t = x_t - \left[\frac{b\phi_1}{(1 - \phi_1)} \right] t$$

Surtout, à partir de (7) vous pouvez aussi montrer que :

$$\begin{aligned} \lim_{t \rightarrow \infty} E[y_t] &= \frac{\phi_0}{(1 - \phi_1)} - b \frac{\phi_1}{(1 - \phi_1)^2}, \text{ et} \\ \lim_{t \rightarrow \infty} Var[y_t] &= \frac{\sigma_u^2}{(1 - \phi_1^2)} \end{aligned}$$

Ainsi, les deux premiers moments de y_t tendent vers ceux d'une variable gouvernée par un AR(1) stationnaire d'écriture :

$$z_t = \left[\phi_0 - b \frac{\phi_1}{(1 - \phi_1)} \right] + \phi_1 z_{t-1} + u_t \quad (9)$$

Les estimateurs de l'équation (9) qui est une réécriture de l'équation initiale (6) convergent donc à la vitesse usuelle \sqrt{T} . En conséquence, on peut appliquer les tests d'hypothèses usuels sur les coefficients après avoir estimé (6) même si elle met en jeu une variable non stationnaire.

Ce résultat signifie que si on peut réécrire le processus avec trend déterministe comme un processus stationnaire sur une variable "*détrendée*", alors on retrouve asymptotiquement sur les estimateurs du processus initial les propriétés des estimateurs OLS. Ce sera le cas par exemple si on considère un AR(p) en lieu et place de l'AR(1) dans (6) comme dans :

$$x_t = \phi_0 + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \dots + \phi_p x_{t-p} + bt + u_t, \quad b \neq 0, \quad u_t \text{ bruit blanc} \quad (10)$$

3. On peut encore montrer qu'un test joint sur a et b posséderait une distribution de Fisher.

Même si x est non stationnaire en raison de la présence du trend bt , on peut estimer ce modèle par OLS et réaliser les tests de significativité des coefficients habituels à condition que les racines du polynôme $(1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p)$ soient toutes de module supérieur à l'unité. De même les OLS s'appliquent si on prend une forme "augmentée" du trend déterministe bt en lui adjoignant un ensemble de variables indicatrices décrivant une saisonnalité additive, comme par exemple sur données trimestrielles, dans :

$$x_t = \phi_0 + \phi_1 x_{t-1} + bt + b_1 D_{1t} + b_2 D_{2t} + b_3 D_{3t} + u_t, \quad |\phi_1| < 1, \quad b \neq 0, b_1 \neq 0, b_2 \neq 0, b_3 \neq 0, \quad (11)$$

avec u_t bruit blanc et $D_{it} = 1$ si t correspond au trimestre i , 0 sinon, pour $i \in \{1, 2, 3\}$

1.3 prévisions dans les processus à trend déterministe

En matière de construction de prévision sur la variable x initiale, cette logique a des implications puisque :

$$\begin{aligned} {}_t x_{t+h} &= E_t[x_{t+h}] \\ &= E_t[T_{t+h}] + E_t[c_{t+h} + u_{t+h}] \\ &= E_t[T_{t+h}] + E_t[\tilde{x}_{t+h}] \\ &= {}_t T_{t+h} + {}_t \tilde{x}_{t+h} \end{aligned}$$

Il suffit donc pour obtenir des prévisions sur la variable non stationnaire x d'ajouter la prévision du trend et la prévision de la composante stationnaire, cette dernière étant connue si le trend lui-même est connu sur la période d'observation. Dans ces conditions, si on suppose que le trend est déterministe, alors on sait le prévoir sans erreur et la seule difficulté est de connaître ce trend sur la période d'observation : il faut résoudre un problème d'estimation, ce qui sera abordé dans la section suivante.

Une autre conséquence de cette hypothèse de trend déterministe est que

$$Var_t[{}_t x_{t+h}] = Var_t[{}_t \tilde{x}_{t+h}] \quad (12)$$

i.e. la variance conditionnelle de la variable non stationnaire est égale à la variance conditionnelle de sa composante stationnaire.

Par ailleurs, comme \tilde{x} est une variable stationnaire, nous connaissons deux propriétés importantes :

1. $\lim_{h \rightarrow \infty} E_t[{}_t \tilde{x}_{t+h}] = E[\tilde{x}_t] = 0$
2. $\lim_{h \rightarrow \infty} Var_t[{}_t \tilde{x}_{t+h}] = Var[\tilde{x}_t] < \infty$

1.3.1 les chocs cycliques et aléatoires n'ont que des effets transitoires

En effet, d'après la première propriété :

$$\lim_{h \rightarrow \infty} {}_t x_{t+h} = \lim_{h \rightarrow \infty} {}_t T_{t+h} + \lim_{h \rightarrow \infty} {}_t \tilde{x}_{t+h} = \lim_{h \rightarrow \infty} {}_t T_{t+h} = T_{t+h} \quad (13)$$

Ainsi, la prévision à un horizon lointain de la variable non stationnaire est égale à la valeur du trend à cet horizon : elle ne dépend pas des valeurs actuelles des composantes cycliques et aléatoires. Cela signifie que, quels que soient les chocs conjoncturels subis aujourd'hui, qu'ils soient favorables ou défavorables, la prévision à long terme de x ne varie pas. Par exemple, sur la figure 1, si on demande en 2005 des prévisions sur x pour 2012, alors que l'on est à un sommet d'un cycle conjoncturel, et que l'on repose la même question en 2006, alors que l'on est à un creux d'un cycle conjoncturel, la réponse sera toujours la même, à savoir x devrait être aux environs de 1200.

Naturellement les chocs qui affectent le trend, qualifiés de structurels, peuvent avoir des effets de long terme.

1.3.2 la largeur des IC afférents aux prévisions de la non stationnaire est finie

Cette proposition découle directement de la deuxième propriété énoncée ci-dessus : la taille des intervalles de confiance que l'on construit pour les prévisions tendent, avec l'horizon de prévision, vers une limite finie qui dépend de la variance de la composante stationnaire de x .

2 Les processus à trend stochastique

Dans les processus à trend déterministe, quelle que soit la fonction définissant ce trend, sa variation est déterministe, *i.e.* pour toute date t et n'importe quel horizon h pris au hasard, on peut prévoir sans se tromper le montant $T_{t+h} - T_t$. Là encore le cas le plus simple correspond à celui du trend linéaire puisque, si $T_t = a + bt$ alors $\Delta T_t = T_t - T_{t-1} = b$ et $T_{t+h} - T_t = bh$.

Une généralisation de cette configuration est de poser que la variation du trend entre deux dates n'est pas déterministe mais est la réalisation d'une aléatoire, soit $\Delta T_t = T_t - T_{t-1} = b_t$ et $Var(b_t) = \sigma_b^2$. On voit immédiatement que si $\sigma_b^2 = 0$, la distribution de l'aléatoire b_t dégénère, b_t devient une constante et on trouve bien que le trend déterministe est un cas particulier de ces nouveaux trends à variation aléatoire que l'on va qualifier logiquement de trends stochastiques.

2.1 processus à racine unitaire et processus intégrés

Le modèle de la marche au hasard est le plus simple des processus à trend stochastique, il pose que l'aléatoire b_t est un bruit blanc et que la variable x n'est composée que d'un trend. Soit donc :

$$x_t = T_t \text{ et } \Delta x_t = u_t \quad (14)$$

où u_t est un bruit blanc de variance σ_u^2 .

- Ce type de processus est qualifié de processus à racine unitaire. Pour comprendre cette terminologie, il suffit de remarquer que $\Delta x_t = u_t \Leftrightarrow (1 - L)x_t = u_t$: c'est effectivement un AR(1) possédant une racine unitaire dans son polynôme caractéristique. En conséquence, on sait déjà qu'il ne sera pas stationnaire, et en effet on montre aisément que $x_t = x_0 + u_1 + u_2 + \dots + u_t$ et donc que $Var[x_t|x_0] = t\sigma_u^2$.
- par ailleurs, une variable est dite intégrée d'ordre d s'il faut la différencier d fois afin d'obtenir une transformée stationnaire. On voit que si x est une marche au hasard, alors elle est intégrée d'ordre 1 puisque $\Delta x_t = u_t$ est stationnaire, ce qu'on notera $x_t \sim I(1)$. Par extension, une variable stationnaire sera dite intégrée d'ordre 0. Par exemple on écrira ici $u_t \sim I(0)$ ou encore $\Delta x_t \sim I(0)$.

2.2 les processus ARIMA

On vient de voir qu'une variable x est $I(d)$ si $y_t = \Delta^d x_t$ est stationnaire. Dans ce cas y_t est susceptible d'obéir à un ARMA(p,q), et plutôt que de faire référence à cette variable intermédiaire y , on dira que x_t est gouvernée par un ARIMA(p,d,q), c'est-à-dire qu'une fois différenciée d fois, la variable non stationnaire x_t obéit à un ARMA(p,q). On aura donc :

$$x_t \sim ARIMA(p, d, q) \Leftrightarrow \Phi(L)\Delta^d x_t = \theta(L)u_t \quad (15)$$

où u_t est comme d'habitude un processus en bruit blanc, et où $\phi(L)$ a toutes ses racines à l'extérieur du cercle complexe unitaire.

Remarques :

- Attention : $\Delta^d = (1 - L)^d \neq (1 - L^d)$. Par exemple, $\Delta^2 x_t = x_t - 2x_{t-1} + x_{t-2} \neq (1 - L^2)x_t = x_t - x_{t-2}$.
- beaucoup de séries économiques semblent être $I(0)$ ou $I(1)$. On peut rencontrer, beaucoup plus rarement des $I(2)$ et apparemment, la probabilité de trouver une série qui devrait être différenciée 3 fois pour en donner une qui soit stationnaire est proche de zéro.
- Ces variables intégrées sont aussi qualifiées de stationnaires en différence, *Difference Stationary*, ou *DS*, pour les distinguer des séries stationnaires autour d'un trend déterministe, ou *Trend Stationary*, *TS*.

2.2.1 processus à racine unitaire, processus intégrés, processus ARIMA

Supposons l'écriture suivante : $\phi(L)x_t = \theta(L)u_t$ où u est un bruit blanc, $\phi()$ et $\theta()$ deux polynômes en L de degrés respectifs p et q . Ce n'est pas obligatoirement un ARMA(p,q) car $\phi(L)$ contient éventuellement une ou plusieurs racines unitaires. Admettons d'ailleurs qu'il en possède une. Dans ce cas, on peut factoriser $\phi(L)$ selon $\phi(L) = (1-L)\phi^*(L)$ avec donc $\phi^*(L)$ un polynôme en L de degré $p-1$. Ainsi, avec une racine unitaire dans $\phi(L)$, on a :

$$\begin{aligned}\phi(L)x_t &= \theta(L)u_t \\ \Leftrightarrow \phi^*(L)(1-L)x_t &= \theta(L)u_t \\ \Leftrightarrow \phi^*(L)\Delta x_t &= \theta(L)u_t\end{aligned}$$

Il n'y a alors plus de racine unitaire dans $\phi^*(L)$ qui est un polynôme de degré $p-1$ en L . Comme on exclut le cas de racine explosive, cela signifie que ses $p-1$ racines sont en module supérieures à l'unité et donc que la dernière égalité est l'équation d'un ARMA($p-1,q$) stationnaire sur Δx_t . En d'autres termes la variable x_t en niveau n'est pas stationnaire puisque son polynôme AR possède une racine unitaire, mais en revanche Δx_t l'est : x_t est intégrée d'ordre 1 : le nombre de racines unitaires, ici 1, dans le processus AR afférent à x_t définit son ordre d'intégration et en conséquence la valeur de d dans la modélisation ARIMA(p,d,q).

Pour vous convaincre de l'identité existant entre le nombre de racines unitaires dans le polynôme autorégressif afférent à la variable en niveau et son ordre d'intégration, supposez que $\phi(L)$ de degré $p \geq 2$ possède deux racines unitaires, il vient : factorisation sur une première racine unitaire :

$$\phi(L) = \phi^*(L)(1-L)$$

et $\phi^*(L)$ est un polynôme de degré $p-1$ qui contient encore une racine unitaire. On factorise :

$$\phi^*(L) = \phi^{**}(L)(1-L)$$

et $\phi^{**}(L)$ ne contient plus que des racines à l'extérieur du cercle complexe unitaire. Au total :

$$\begin{aligned}\Phi(L)x_t &= \theta(L)u_t \\ \Leftrightarrow \phi^*(L)(1-L)x_t &= \theta(L)u_t \\ \Leftrightarrow \phi^{**}(L)(1-L)^2x_t &= \theta(L)u_t \\ \Leftrightarrow \phi^{**}(L)\Delta^2x_t &= \theta(L)u_t\end{aligned}$$

et la dernière égalité est l'écriture d'un ARMA($p-2,q$) sur la variable stationnaire Δ^2x_t , *i.e.* $x_t \sim \text{ARIMA}(p-2,2,q)$. Il a donc été nécessaire de différencier deux fois la variable x_t pour la stationnariser : elle est I(2). En d'autres termes, chaque différence première opérée sur une variable supprime une éventuelle racine unitaire présente dans le polynôme AR afférent à la variable en niveau.

On peut se demander alors quelles sont les conséquences d'une différenciation inutile, par exemple de passer à la différence première d'une variable dont le polynôme AR ne possède pas de racine unitaire, c'est ce qu'on appelle la sur-différenciation : Soit $\phi(L)x_t = \theta(L)u_t$ un processus ARMA(p,q) sur la variable stationnaire x_t . Le passage aux différences premières mène à la variable Δx_t , or on a : $\phi(L)(1-L)x_t = \theta(L)(1-L)u_t$, soit encore $\phi(L)\Delta x_t = \theta^*(L)u_t$ avec $\theta^*(L) = \theta(L)(1-L)$: la différenciation inutile introduit une racine unitaire dans la composante MA de sorte que la série sur-différenciée Δx_t obéit à un ARMA($p,q+1$) non inversible.

2.3 comparaison trend déterministe - trend stochastique : AR(1)-TS versus AR(1)-DS

Pour comparer et illustrer les différences existantes entre deux types de processus non stationnaires, nous allons considérer un processus simple de type AR(1) dans chacune des classes :

1. un AR(1)-TS, somme d'un trend déterministe linéaire et d'un AR(1) stationnaire :

$$x_t = a + bt + u_t, \tag{16}$$

$$u_t = \phi u_{t-1} + \epsilon_t \tag{17}$$

où ϵ_t est un processus en bruit blanc, $|\phi| < 1$ et $b \neq 0$.

2. un AR(1)-DS, somme d'un trend déterministe linéaire et d'un trend stochastique (une marche au hasard) :

$$y_t = a + bt + v_t, \quad (18)$$

$$v_t = v_{t-1} + \epsilon_t \quad (19)$$

Les deux processus sont non stationnaires puisque $E[x_t] = a + bt$. Cependant, si x_t a une variance bien définie, ce n'est pas le cas de y_t :

$$\text{Var}[x_t] = \frac{\sigma_\epsilon^2}{1 - \phi_2},$$

$$\text{Var}[y_t] = t \sigma^2 \epsilon$$

En conséquence, les largeurs des intervalles de confiance construits sur les prévisions vont augmenter dans les deux cas avec l'horizon de prévision mais alors que pour x elles atteindront une limite dépendante de la variance non conditionnelle finie, pour y cette largeur tend vers l'infini.

Toutefois la différence la plus importante entre les deux processus provient de l'impact des chocs ϵ sur chacune des variables. En effet, si on suppose que le choc d'innovation n'affecte pas la pente du trend déterministe de long terme alors :

$$\frac{\delta x_t}{\delta \epsilon_{t-i}} = \phi^i, \text{ et, comme } |\phi| < 1, \lim_{i \rightarrow \infty} \phi^i = 0. \text{ en revanche :}$$

$$\frac{\delta y_t}{\delta \epsilon_{t-i}} = 1.$$

Alors que l'impact d'un choc passé finit par disparaître pour le processus AR(1)-TS, il perdure pour le processus AR(1)-DS qui se caractérise donc par ce que l'on nomme l'hystérèse des chocs.

Les figures 2, 3 et 4 illustrent ces différences entre séries TS et séries DS. La première représente deux séries simulées à partir des équations (16) et (17) pour l'une (AR1-TS), et des équations (18) et (19) pour l'autre (AR1-DS). Pour cette simulation on a fixé $a = 10$, $b = 0.05$, $\phi = 0.8$ et ϵ_t sont des tirages de pseudo-gaussiennes centrées réduites. Le même ensemble de résidus ϵ a été utilisé pour construire les deux séries de 172 observations. On peut déjà noter la forte similitude des profils des deux trajectoires : si on ne disposait que de leur seul aspect visuel il serait pratiquement impossible de savoir laquelle est TS et laquelle est DS.

Nous avons ensuite imposé un choc négatif en août 2010 : $\epsilon_{2010/8} = -5$. L'ampleur de ce choc, égal à 5 fois l'écart-type du résidu simulé, doit permettre de mettre en évidence les différences de réaction des deux séries. Il s'agit d'un choc unique : passé cette date, les valeurs des ϵ_t reprennent leurs valeurs initiales sur la période septembre 2010 - avril 2018. On peut observer les conséquences du choc sur la série à trend déterministe dans la figure 3. On voit qu'un écart important se creuse entre la trajectoire initiale et celle qui suit le mois de la perturbation, mais cet écart, important en 2010 et sur la première moitié de 2011 tend à s'annuler dans les derniers mois de cette année 2011. Dès 2012, les deux trajectoires, avec et sans choc, ne se distinguent plus, ce qui témoigne bien d'un impact seulement transitoire des perturbations dans les modèles TS. La figure 4 reproduit trajectoires de la série DS. L'impact du choc négatif apparaît également clairement en août 2010 mais on note surtout l'absence de retour de la trajectoire perturbée vers les valeurs initiales : la nouvelle trajectoire se retrouve de façon permanente en-dessous de l'initiale illustrant donc cette permanence de l'impact des chocs dans les séries DS⁴.

2.4 mesure de l'impact d'un choc d'innovation sur une série I(1)

On vient de montrer qu'un choc d'innovation qui n'affecte pas l'éventuel trend déterministe présent dans une série a un impact nul à long terme si la série est de type DS et non nul si elle est de type DS. Nous allons voir que dans ce dernier cas il est possible de préciser l'amplitude de l'effet sur un horizon quelconque fini et sur un horizon infini. Pour rester dans un cadre simple, nous allons considérer une série I(1) sans constante ni trend déterministe, sachant que si le choc d'innovation n'affecte pas ces dernières composantes, leur présence ou leur absence n'affecte en rien son impact sur

4. Naturellement les hypothèses sous lesquelles cette illustration est réalisée sont essentielles pour comprendre les résultats obtenus : dans le cas de la série TS les chocs d'innovation ϵ_t sont supposés ne pas affecter le trend et notamment sa pente donnée par le coefficient constant b . En revanche dans la modélisation DS, comme $\Delta y_t = b + \epsilon_t$, la pente du trend varie en tout point du temps selon les valeurs de ϵ_t . Pour cette raison on qualifie d'ailleurs alors la tendance de *local trend*

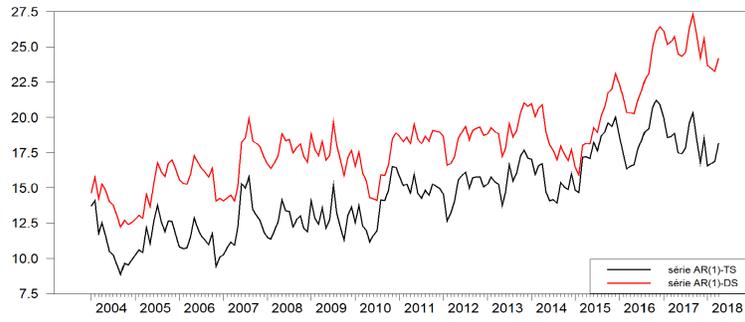


FIGURE 2 – Exemple de trajectoires de séries à trend déterministe et à trend stochastique

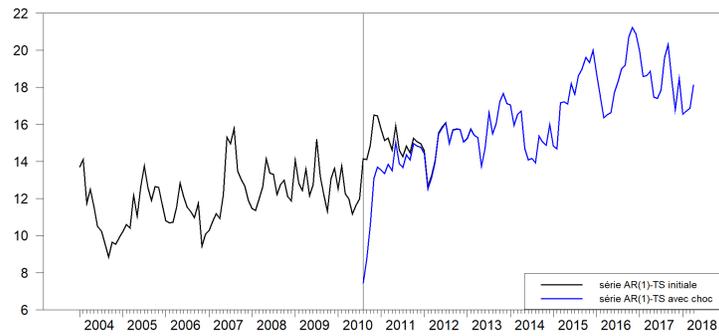


FIGURE 3 – Exemple de l'impact d'un choc ponctuel sur la trajectoire d'une série TS

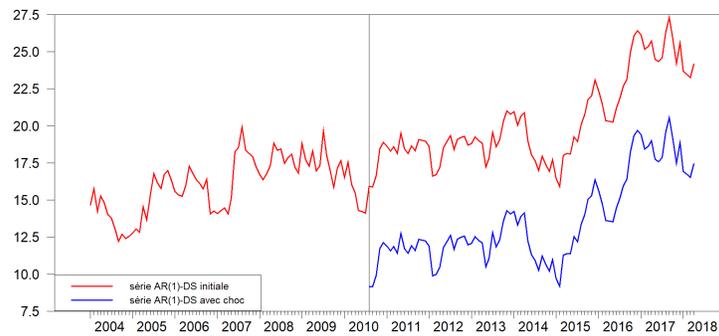


FIGURE 4 – Exemple de l'impact d'un choc ponctuel sur la trajectoire d'une série DS

le niveau de la série elle-même.

Soit donc une variable y_t . On a évidemment :

$$\begin{aligned} y_{t+1} &= y_t + (y_{t+1} - y_t) = y_t + \Delta y_{t+1} \\ y_{t+2} &= y_t + (y_{t+1} - y_t) + (y_{t+2} - y_{t+1}) = y_t + \Delta y_{t+1} + \Delta y_{t+2} \\ &\vdots \\ y_{t+h} &= y_t + (y_{t+1} - y_t) + (y_{t+2} - y_{t+1}) + \dots + (y_{t+h} - y_{t+h-1}) = y_t + \Delta y_{t+1} + \Delta y_{t+2} + \dots + \Delta y_{t+h} \end{aligned}$$

Si cette variable est intégrée d'ordre 1 alors par définition quel que soit l'entier i , Δy_{t+i} est $I(0)$ et possède une écriture de Wold de la forme :

$$\Delta y_{t+i} = \sum_{j=0}^{\infty} \psi_j u_{t+i-j} = \psi(L)u_t \text{ avec } \psi_0 = 1$$

où u est un bruit blanc. En combinant les deux dernières égalités il vient :

$$y_{t+h} = y_t + \sum_{j=0}^{\infty} \psi_j u_{t+1-j} + \sum_{j=0}^{\infty} \psi_j u_{t+2-j} + \dots + \sum_{j=0}^{\infty} \psi_j u_{t+h-j}$$

On étudie maintenant la réponse de cette variable y au choc de $t+1$. On voit aisément que :

$$\begin{aligned} \frac{\delta y_{t+1}}{\delta u_{t+1}} &= \psi_0 = 1 \\ \frac{\delta y_{t+2}}{\delta u_{t+1}} &= \psi_0 + \psi_1 \\ &\vdots \\ \frac{\delta y_{t+h}}{\delta u_{t+1}} &= \psi_0 + \psi_1 + \dots + \psi_{h-1}, \text{ et enfin} \\ \lim_{h \rightarrow \infty} \frac{\delta y_{t+h}}{\delta u_{t+1}} &= \sum_{j=0}^{\infty} \psi_j = \psi(1) \end{aligned}$$

Ce dernier ensemble d'égalités donne donc la mesure de l'impact sur le niveau de notre variable du choc d'innovation qui se produit en $t+1$: à cette date l'effet est naturellement égal à l'ampleur de la perturbation, par la suite cet effet pour un horizon quelconque dépend des coefficients de l'écriture de Wold afférente à Δy_t et ce qu'on nomme l'effet de long terme, *i.e.* l'impact évalué sur un horizon infini est simplement donné par la somme de tous ces coefficients. Ainsi, si $\psi(1) > 1$ alors on a un effet d'amplification au sens où l'amplitude de l'impact à long terme de la perturbation est supérieure à son amplitude initiale et à l'inverse, on observe un effet d'amortissement à long terme lorsque la valeur absolue de $\psi(1)$ est inférieure à l'unité.

3 Les tests de racine unitaire

En pratique, le simple graphique d'une trajectoire aide naturellement à prendre une décision quant à la stationnarité ou à la non stationnarité d'une variable, même s'il peut apparaître des situations dans lesquelles la décision à prendre n'est pas évidente. Un signal peut aussi être envoyé par le corrélogramme : de fortes autocorrélations positives associées à une décroissance lente de la fonction d'autocorrélation estimée sont interprétées comme des signes de non stationnarité. La raison en est simple : si une série possède une tendance, alors une fois qu'une observation est passée au-dessus de la moyenne empirique (pour une tendance haussière) ou en-dessous (pour une tendance baissière), les observations qui suivent vont plutôt se trouver dans la même situation et les écarts à la moyenne seront en majorité de même signe. En conséquence la somme des produits des écarts à la moyenne va avoir tendance à croître, poussant à la hausse les autocovariances. Dans cette configuration, il est nécessaire de considérer des retards relativement importants pour trouver des écarts à la moyenne ayant des signes différents qui devraient alors commencer à limiter la hausse, puis provoquer la baisse des autocovariances.

Par ailleurs, si on a admis qu'une variable est non stationnaire en espérance, alors une autre décision doit également être prise. Que ce soit pour discuter de l'impact des perturbations et les mesures à prendre pour y faire face, ou pour construire des prévisions sachant qu'alors il y a une étape nécessaire de stationnarisation, soit par régression sur un

trend déterministe, soit par différenciation pour les processus à trend stochastique, il est important de savoir quel type de tendance est éventuellement présent dans une variable. Le premier test permettant de s'attaquer à ce problème a été proposé par Dickey et Fuller⁵. Ce premier test, qui possède trois versions, est adapté au seul processus AR(1). Après les avoir présentées, nous verrons par la suite comment il a été étendu aux processus AR(p) et ARMA(p,q). Son hypothèse nulle est qu'une racine unitaire est présente dans le polynôme AR afférent à la variable de travail, *i.e.* que la variable est intégrée d'ordre 1.

3.1 les tests DF de présence d'une racine unitaire dans le polynôme caractéristique d'un AR(1)

3.1.1 Les statistiques τ , τ_μ et τ_t de Dickey-Fuller

Le test DF est proposé en trois versions selon que le processus intègre ou non une constante ou un trend déterministe.

1. le test τ : le processus considéré est un AR(1) sans composante déterministe :

$$x_t = \phi_1 x_{t-1} + u_t \quad (20)$$

Comme on exclut les processus explosifs⁶, on va opposer :

- H0 : $\phi_1 = 1$: il existe une racine unitaire, à
- H1 : $|\phi_1| < 1$: le processus est stationnaire

Clairement la formulation de H1 n'est pas adaptée à la conduite usuelle d'un test sur un coefficient car il ne correspond ni à une formulation bilatérale ni à une formulation unilatérale de l'alternative. On peut toutefois échapper aisément à cette difficulté : Dickey et Fuller proposent de réécrire (20) de façon équivalente, en soustrayant x_{t-1} à gauche et à droite, comme

$$\begin{aligned} \Delta x_t &= (\phi_1 - 1)x_{t-1} + u_t \\ &= b x_{t-1} + u_t \end{aligned} \quad (21)$$

Avec cette réécriture, on va alors opposer :

- H0 : $b = 0$: il existe une racine unitaire, à
- H1 : $b < 0$: le processus est stationnaire

et on est revenu à une formulation unilatérale de l'alternative pour un test de nullité d'un coefficient. Dickey et Fuller proposent alors d'estimer (21) par OLS et de construire la stat $\hat{b}/s_{\hat{b}}$, *i.e.* de calculer le student correspondant à un test de nullité. La difficulté vient de ce que sous H0 on ne retrouve pas les propriétés usuelles des OLS : la statistique ne possède pas une distribution de student⁷. Pour insister sur ce point, Dickey-Fuller nomment la statistique en question τ : elle ressemble par construction à un student, mais ce n'en est pas un.

2. le test τ_μ : le processus considéré est un AR(1) avec terme constant :

$$x_t = \phi_0 + \phi_1 x_{t-1} + u_t \quad (22)$$

Cette version vise à donner plus de généralité au test de racine unitaire. En effet dans la version τ , sous l'alternative, x est un AR stationnaire autour d'une espérance nulle. Dans ces conditions, il ne peut pas s'appliquer à des variables à valeurs toujours positives tels que taux de change, ventes et/ou achats de biens en volume et en valeur, prix, etc... Une solution est d'autoriser que sous l'alternative la série fluctue autour d'une constante, strictement positive (ou négative). On sait alors qu'avec (20), si $|\phi_1| < 1$ alors $E[x_t] = \phi_0/(1 - \phi_1)$ qui peut être positive ou négative selon le signe de ϕ_0 . Dans ce cadre d'application plus grand, la question reste toujours la même, à savoir existe-t'il une racine unitaire dans la série x ? En excluant toujours les séries explosives, on opposera donc encore :

5. Dickey, D. A. ; Fuller, W. A. (1979). Distribution of the Estimators for Autoregressive Time Series with a Unit Root, *Journal of the American Statistical Association* 74, 427-431.

6. Ceux-ci se caractérisent par un coefficient $|\phi_1|$ et prennent des valeurs allant très vite vers l'infini ce que vous pouvez vérifier aisément par simulation. Pour cette raison on les considère comme non pertinents pour la modélisation des séries économiques.

7. On peut aisément illustrer la survenue d'une difficulté pour la réalisation du test usuel : soit $x_t = \phi_1 x_{t-1} + u_t$ avec u gaussien, on a $(\hat{\phi}_1 - \phi_1) \sim \mathcal{N}(0, \sigma_u^2 (\sum_{t=2}^T x_{t-1}^2)^{-1})$, soit encore $\sqrt{T}(\hat{\phi}_1 - \phi_1) \sim \mathcal{N}(0, \sigma_u^2 (\frac{1}{T} \sum_{t=2}^T x_{t-1}^2)^{-1})$. Comme le processus est centré, la quantité $\frac{1}{T} \sum_{t=2}^T x_{t-1}^2$ tend vers la variance de x , γ_0 . Or, pour un AR(1) stationnaire, $\gamma_0 = \frac{\sigma_u^2}{1 - \phi_1^2}$. Au final, il vient $\sqrt{T}(\hat{\phi}_1 - \phi_1) \sim \mathcal{N}(0, 1 - \phi_1^2)$. En conséquence, sous l'hypothèse nulle qui nous intéresse ici, la loi de $\sqrt{T}(\hat{\phi}_1 - \phi_1)$ dégénère puisque sous H0 : $\phi_1 = 1$, sa variance s'annule!

- H0 : $\phi_1 = 1$: il existe une racine unitaire, à
- H1 : $|\phi_1| < 1$: le processus est stationnaire

Pour la même raison que précédemment, on va passer à la réécriture de Dickey-Fuller :

$$\begin{aligned}\Delta x_t &= \phi_0 + (\phi_1 - 1)x_{t-1} + u_t \\ &= \phi_0 + bx_{t-1} + u_t\end{aligned}\tag{23}$$

qui conduit au test :

- H0 : $b = 0$: il existe une racine unitaire, versus
- H1 : $b < 0$: le processus est stationnaire

Pour le réaliser, on va estimer (23) et construire la statistique appelée maintenant τ_μ , pour rappeler qu'elle est issue d'une régression avec constante selon : $\tau_\mu = \hat{b}/s_{\hat{b}}$. C'est donc la valeur du student qui testerait la non nullité de b , mais, comme précédemment, sous H0 on ne retrouve pas les lois usuelles des OLS : τ_μ possède une distribution qui n'est pas de student.

3. le test τ_t : le processus considéré est un AR(1) avec terme constant et trend linéaire déterministe. Le test est donc maintenant construit en considérant le processus

$$x_t = \phi_0 + \phi_1 x_{t-1} + \phi_2 t + u_t\tag{24}$$

et il porte toujours sur :

- H0 : $\phi_1 = 1$: il existe une racine unitaire, versus
- H1 : $|\phi_1| < 1$: le processus est stationnaire autour d'un trend linéaire déterministe.

On peut remarquer qu'ici la variable x n'est pas stationnaire, que ce soit sous H0 ou sous H1. L'écriture DF de (24) devient :

$$\begin{aligned}\Delta x_t &= \phi_0 + (\phi_1 - 1)x_{t-1} + \phi_2 t + u_t \\ &= \phi_0 + bx_{t-1} + \phi_2 t + u_t\end{aligned}\tag{25}$$

et le test correspondant reste :

- H0 : $b = 0$: il existe une racine unitaire, versus
- H1 : $b < 0$: le processus est stationnaire

On construit alors la statistique $\tau_t = \hat{b}/s_{\hat{b}}$ qui sous H0 ne suit toujours pas un student et possède une distribution décalée à gauche relativement à celle de τ_μ .

Cette troisième version est supposée corriger une incohérence de la deuxième. En effet, si on reprend le processus avec constante duquel est extrait τ_μ , on voit aisément que :

- Si H0 est vrai, le modèle devient : $x_t = \phi_0 + x_{t-1} + u_t$, soit $\Delta x_t = \phi_0 + u_t$. Ainsi, Δx_t est défini sur l'aléatoire u_t , c'est-à-dire qu'il existe un trend stochastique, ce que l'on veut évidemment sous H0, mais il est aussi défini sur la constante ϕ_0 . Or, si la différence première d'une variable contient une constante alors le niveau de cette variable contient un trend déterministe dont la pente est la constante en question. Au total, sous H0, la variable x_t se caractérise par un trend stochastique local de pente u_t et un trend linéaire déterministe de pente ϕ_0 .
- Si H0 est faux, le modèle qui gouverne x_t est alors $x_t = \phi_0 + \phi_1 x_{t-1} + u_t$ avec $|\phi_1| < 1$: x est une stationnaire qui oscille autour de la constante $\phi_0/(1-\phi_1)$. Elle ne contient ni trend stochastique, ni trend déterministe.

En résumé, dans la deuxième version du test, sous H0 la variable x possède deux trends, un stochastique et un déterministe, sous H1 elle n'en possède plus aucun. Le test τ_μ est donc en réalité un test joint et quand on rejette un test joint, on ne connaît pas la cause du rejet. Par exemple, ici le rejet peut être dû tout autant à l'absence d'un trend déterministe qu'à l'absence d'un trend stochastique. En conséquence, ce test τ_μ ne permet pas de répondre à

l'interrogation initiale qui est de savoir si x contient un trend stochastique.

Compte-tenu de ce constat, l'idée de la troisième version est d'introduire d'office un trend déterministe, $\phi_2 t$, dans le modèle de sorte que sous H0 on aurait les deux trends mais que sous H1 seul disparaîtrait le trend stochastique. Il est aisé de montrer que cette astuce ne fait que repousser la difficulté : avec (24) sous H0 on aurait dans le niveau de x un trend stochastique et un trend quadratique, et sous H1 un trend linéaire déterministe, *i.e.* au final on est toujours sur un test joint qui n'oppose pas simplement présence versus absence d'un trend stochastique. On pourrait évidemment continuer la démarche et introduire un trend quadratique déterministe dans l'écriture initiale. Comme on peut s'en douter, cela ne ferait encore que repousser le problème sans le résoudre.

Au final, et malgré leurs insuffisances, les trois versions, τ , τ_μ , et τ_t du test DF sont toujours communément utilisées⁸.

3.1.2 La recherche des valeurs critiques des tests DF

Nous avons vu que quelle que soit la version du test DF, la statistique est construite comme le serait un student pour effectuer un test de nullité d'un coefficient mais que pour autant elle ne possède pas cette distribution.

La cause de ce résultat est que dans la régression de Dickey-Fuller sous hypothèse de présence d'une seule racine unitaire, l'expliquée qui est Δx_t est I(0) alors que l'explicative, $x_t - 1$ est une I(1) non stationnaire. Cependant, malgré cette non connaissance de la distribution de leurs statistiques, Dickey et Fuller vont pouvoir au moyen de simulations afficher des valeurs critiques pour des seuils de risque usuels. Leur démarche est simple et vous devriez d'ailleurs être en mesure de la répéter.

Le point de départ est comme toujours de construire le test sous H0, *i.e.* en supposant que l'hypothèse nulle est vraie. Dans ce cas, si on prend la première version, on sait que les réalisations de la variable x sont générées par l'équation :

$$x_t = x_{t-1} + u_t \quad (26)$$

Deux premières étapes vont alors se succéder et se répéter :

1. simulation d'une trajectoire pour x d'une longueur T fixée a priori, par exemple de 50 ou 100 observations⁹ et de pseudo-réalisations indépendantes pour u_t possédant une loi désirée, ici gaussienne.
2. réalisation sur cette trajectoire d'une estimation OLS de l'équation de Dickey-Fuller du type
 - $\Delta x_t = b x_{t-1} + \epsilon_t$ de laquelle on tire une statistique τ , ou du type
 - $\Delta x_t = \phi_0 + b x_{t-1} + \epsilon_t$ de laquelle on tire une statistique τ_μ , ou du type
 - $\Delta x_t = \phi_0 + b x_{t-1} + \epsilon_t + \phi_2 t$ de laquelle on tire une statistique τ_t

Les étapes 1 et 2 sont répétées n fois, par exemple 1000 fois. A la fin on dispose donc de 3 échantillons de taille n pour respectivement τ , τ_μ et τ_t . Enfin, la troisième et dernière étape fournit les estimations des valeurs critiques via certains centiles pris dans chacun de ces échantillons :

3. le 95^{ème} centile par exemple donne une estimation de la valeur critique à 5% de risque. En effet, par définition, 5% des valeurs de la statistique construites sous H0 lui sont inférieures, et donc son emploi comme valeur critique mène au rejet de la nulle dans 5% des cas alors qu'elle est vraie, ce qui est bien la définition du risque de première espèce à 5%. Pour les valeurs critiques à 10% et 1%, on prendra simplement respectivement les 90^{ème} et 99^{ème} centiles.

A l'issue de leur travail, Dickey et Fuller observent que les valeurs critiques de ces nouvelles statistiques τ , τ_μ et τ_t sont toutes inférieures à celles d'un student de sorte que l'emploi des valeurs critiques tirées de la student conduirait pour un seuil de risque donné, à rejeter trop souvent l'hypothèse nulle, *i.e.* à admettre trop souvent et à tort l'absence de racine unitaire.

Ils remarquent également que celles de τ_μ sont inférieures à celles de τ et que celles de τ_t sont également inférieures à celles de τ_μ . Ainsi la présence ou l'absence des termes de nuisance que sont la constante ϕ_0 et la pente du trend déterministe, ϕ_2 , affecterait la distribution de la statistique DF¹⁰.

8. Dickey et Fuller ont dans le même travail proposé une autre statistique nommée *rho* pour le même test de présence d'une racine unitaire définie par $\rho = T(\hat{\phi}_1 - 1)$. Notez que bien que celle-ci ait pu sembler être plus puissante que la statistique τ , c'est cette dernière qui a été et est retenue dans pratiquement tous les travaux empiriques qui ont suivi.

9. Ici cela suppose que l'on se donne une valeur initiale, par exemple $x_0 = 0$ et qu'un certain nombre des valeurs initialement générées pour x ne sont pas sauvegardées précisément pour éliminer l'impact de la fixation arbitraire de la condition initiale, c'est ce que l'on nomme le *burn-in*.

10. Par exemple, pour une taille d'échantillon de 600 observations, les valeurs critiques à 5% sont de -1.64 pour un student, de -1.94 pour τ , de -2.86 pour τ_μ et de -3.41 pour τ_t . Notez aussi que le test étant $b = 0$ versus $b < 0$, ces valeurs critiques sont obligatoirement négatives.

Pour clore cette section, remarquez qu'à l'issue de ce travail de Dickey et Fuller au moins une difficulté subsiste : les valeurs critiques qu'ils publient ne concernent que quelques tailles d'échantillon, par exemple $T = 50, 100$ ou 150 : il faut évidemment les adapter, et donc pour chaque étude empirique construire les tables adaptées au nombre d'observations de la série de travail. Fort heureusement, McKinnon a estimé et publié¹¹ des équations simples permettant le calcul de ces bornes critiques en fonction de la taille T

Notez que pour que les distributions de ces statistiques soient celles de Dickey-Fuller, il est essentiel que les résidus des équations (21), (23) et (25) soient des bruits blancs.

3.2 le test ADF : racine unitaire dans le polynôme AR d'un AR(p) ou d'un ARMA(p,q), $p > 0$

La question posée est celle de la présence d'un trend stochastique dans le niveau d'une variable non stationnaire. Elle ne concerne donc pas les processus MA(q) qui sont toujours stationnaires, mais seulement les AR(p) et les ARMA(p,q) avec $p > 0$. On va tout d'abord considérer le test de présence d'une première racine unitaire dans les polynômes autorégressifs de ces processus. Ensuite, comme il s'agit de polynômes de degré pouvant être supérieur à un, et donc ayant plusieurs racines, nous envisagerons le test de présence d'une autre racine unitaire sachant qu'on en a déjà détecté au moins une.

3.2.1 test de présence d'une racine unitaire

- le cas des AR(p) avec $p < \infty$: la proposition essentielle est que dans la réécriture de Dickey-Fuller du processus AR, le coefficient du niveau retardé vaut zéro si une des racines du polynôme $\phi(L)$ est égale à 1. C'est vrai pour le processus AR(1) comme on vient de le voir avec le test DF, et on peut le vérifier aisément sur un AR(2) : si on passe de l'écriture canonique AR(2) à l'écriture de Dickey-Fuller il vient :

$$\begin{aligned}
(1 - \phi_1 L - \phi_2 L^2)x_t &= u_t \\
\Leftrightarrow x_t &= \phi_1 x_{t-1} + \phi_2 x_{t-2} + u_t \\
\Leftrightarrow x_t &= (\phi_1 + \phi_2)x_{t-1} - \phi_2 x_{t-1} + \phi_2 x_{t-2} + u_t \\
\Leftrightarrow x_t &= (\phi_1 + \phi_2)x_{t-1} - \phi_2 \Delta x_{t-1} + u_t \\
\Leftrightarrow \Delta x_t &= (\phi_1 + \phi_2 - 1)x_{t-1} - \phi_2 \Delta x_{t-1} + u_t \\
\Leftrightarrow \Delta x_t &= b x_{t-1} - \phi_2 \Delta x_{t-1} + u_t
\end{aligned} \tag{27}$$

où $b = -\phi(1)$ et évidemment s'il y a une racine unitaire $b = 0$. De même pour un AR(3) :

$$\begin{aligned}
(1 - \phi_1 L - \phi_2 L^2 - \phi_3 L^3)x_t &= u_t \\
x_t &= \phi_1 x_{t-1} + \phi_2 x_{t-2} + \phi_3 x_{t-3} + u_t \\
\Leftrightarrow x_t &= (\phi_1 + \phi_2 + \phi_3)x_{t-1} - (\phi_2 + \phi_3)(x_{t-1} - x_{t-2}) - \phi_3(x_{t-2} - x_{t-3}) + u_t \\
\Leftrightarrow \Delta x_t &= (\phi_1 + \phi_2 + \phi_3 - 1)x_{t-1} - (\phi_2 + \phi_3)\Delta x_{t-1} - \phi_3 \Delta x_{t-2} + u_t \\
\Leftrightarrow \Delta x_t &= b x_{t-1} - \phi_2^* \Delta x_{t-1} - \phi_3 \Delta x_{t-2} + u_t
\end{aligned} \tag{28}$$

Avec encore en cas de racine unitaire $b = -\phi(1) = 0$. Plus généralement, pour un AR(p) on aura :

$$\begin{aligned}
x_t &= \phi_0 + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \dots + \phi_p x_{t-p} + u_t \\
\Leftrightarrow \Delta x_t &= b x_{t-1} + \sum_{i=1}^{p-1} \phi_i^* \Delta x_{t-i} + u_t
\end{aligned} \tag{29}$$

avec toujours $b = -\phi(1)$.

Dans cette équation (29), les termes en Δx retardés s'appellent des termes d'augmentation, d'où la dénomination du test : ADF pour *Augmented Dickey-Fuller*. Ils sont absolument nécessaires pour assurer que les résidus des régressions soient orthogonaux entre eux et donc que les statistiques de Dickey-Fuller possèdent bien les distributions attendues.

On peut concevoir (29) comme étant une généralisation du test DF réalisé sans constante ni trend. Naturellement

11. MacKinnon, J. G. (1991), Critical values for cointegration tests, Chapter 13 in *Long-Run Economic Relationships : Readings in Cointegration*, ed. R. F. Engle and C. W. J. Granger. Oxford, Oxford University Press ; et plus récemment, MacKinnon J. G., Critical Values for Cointegration Tests, *Queen's Economics Department Working Paper No. 1227*, 2010.

les deux autres versions peuvent également être généralisées et le test *ADF* pourra donc être effectué au moyen de la régression (29) mais encore avec les versions suivantes :

$$\Delta x_t = \phi_0 + bx_{t-1} + \sum_{i=1}^{p-1} \phi_i^* \Delta x_{t-i} + u_t, \text{ où on a ajouté une constante,} \quad (30)$$

et

$$\Delta x_t = \phi_0 + dt + bx_{t-1} + \sum_{i=1}^{p-1} \phi_i^* \Delta x_{t-i} + u_t, \text{ où on a ajouté une constante et un trend.} \quad (31)$$

les adjonctions de ces termes étant justifiées par les mêmes raisons que celles vues dans la présentation du test *DF*. Il suffira donc d'estimer par OLS des équations du type (29), (30) ou (31). On récupérera ainsi respectivement les statistiques $\tau = \hat{b}/s$, $\tau_\mu = \hat{b}/s$ et $\tau_t = \hat{b}/s$. L'important ici est que ces statistiques ont sous H_0 la même distribution que les statistiques de même nom construites dans le cadre du test *DF*. L'intuition de ce résultat vient d'une part que l'ajout des variables Δx retardées permet de retrouver l'hypothèse de bruit blanc sur le résidu des régressions et d'autre part qu'en présence d'une seule racine unitaire les termes d'augmentation sont $I(0)$ et n'ont de ce fait pas d'impact sur la distribution de la statistique qui serait obtenue en leur absence.

A ce stade, vous devez comprendre que le choix du nombre d'augmentations est un point essentiel pour la mise en oeuvre du test *ADF* : si ce nombre est insuffisant il subsistera de l'autocorrélation dans la série résiduelle u_t et les statistiques *ADF* ne posséderont pas leur distribution asymptotique, s'il est trop important la présence de variables non pertinentes va affecter la puissance des tests. Ng et Perron ont proposé une procédure de réduction : l'utilisateur ajuste une régression du type (29), (30) ou (31) en imposant un nombre *pmax* de termes Δx retardés tel que le test de Ljung-Box ne détecte plus d'autocorrélation sérielle dans les résidus. Il teste ensuite par un student la nullité du coefficient de la dernière augmentation. Si celle-ci est non significative, il réduit *pmax* d'une unité, vérifie l'absence d'autocorrélation résiduelle et continue la réduction jusqu'à obtenir une régression dans laquelle le dernier terme d'augmentation est significatif et que Ljung-Box ne signale pas la présence de corrélations non nulles dans les résidus. Si on note \hat{p} le nombre d'augmentations alors atteint, la procédure s'achève avec un test *ADF* qui est réalisé avec ce nombre \hat{p} de retards. Schwert a de son côté proposé une règle de détermination de la valeur initiale *pmax* qui la fait dépendre de la taille de l'échantillon : $pmax = 12 \left(\frac{T}{100}\right)^{1/4}$. D'autres auteurs utilisent des critères de sélection pour trouver la régression de type optimale parmi (23), (24) ou (25) pour le test *DF*. L'important pour la validité du test *ADF* comme d'ailleurs du test *DF* étant de vérifier l'absence de corrélation sérielle dans la régression retenue pour le réaliser.

L'autre difficulté est évidemment celle du choix de la régression à effectuer : sans constante, avec constante, avec constante et trend. Dans les premiers travaux ce choix était gouverné par une succession de tests. Par exemple, partant du modèle le plus général (31), on teste $H_0 : b = 0$ au moyen de τ_t . En cas de rejet, on effectue un test standard¹² de nullité de d qui, si elle est rejetée, conduit à retenir un processus TS sur la variable d'intérêt et sinon conduit à un processus stationnaire autour d'une constante non nulle. Si au contraire on ne rejette pas la nullité de b alors on ajuste un modèle en différence avec dérive tel que¹³ $\Delta x_t = \phi_0 + u_t$. Un test de $H_0 : \phi_0 = 0$ permettant de retenir un processus de marche au hasard en cas de non rejet, et de marche au hasard avec dérive sinon¹⁴. La difficulté avec cette démarche est qu'avec la succession de tests on ne maîtrise pas le seuil de risque avec lequel le processus terminal est sélectionné. Dans les travaux actuels on préfère souvent introduire un a-priori plus ou moins justifié permettant de retenir une modélisation alternative raisonnable en cas de rejet de $H_0 : b = 0$. Par exemple, si la série de travail est à valeurs positives, alors on sait que l'alternative ne peut pas être un processus stationnaire d'espérance nulle : l'équation (29) qui impliquerait alors un processus stationnaire autour de zéro est inadaptée et donc la statistique τ est délaissée au profit de τ_μ ou τ_t . Si le simple graphique de cette série fait apparaître une tendance dans son niveau moyen, alors c'est également τ_μ qui est abandonnée car son alternative en cas de non nullité de b est un processus stationnaire autour d'une constante. Enfin en l'absence d'a-priori certains ont recours aux critères de sélection pour choisir entre (29), (30) et (31), l'idée est que l'on est en présence de modèles concurrents parmi lesquels il faut sélectionner un modèle optimal. Il semble alors que le critère de Schwarz doit être le plus

12. Qui est alors valide comme nous l'avons vu précédemment.

13. Pour simplifier on a raisonné en l'absence de termes d'augmentation mais leur présence ne changerait rien au raisonnement qui est mené.

14. Naturellement si des augmentations sont présentes le processus final n'est pas une marche au hasard, mais une équation où la différence première est expliquée par des différences premières retardées avec ou sans constante selon le cas.

souvent préféré à celui d'Akaike et ses variantes ¹⁵.

- le cas des ARMA(p,q) à composante MA inversible.

Il est réglé très simplement en suivant un raisonnement proposé par Saïd et Dickey. L'idée est que ces ARMA ont une écriture AR(∞) équivalente. On cherche donc une approximation de cet AR, de préférence avec le critère d'Akaike et on reproduit à partir de cette première approximation la démarche que l'on vient juste d'exposer pour les AR(p). La justification est que si le vrai processus est non stationnaire alors l'approximation doit aussi l'être et que si le vrai processus est stationnaire alors l'approximation doit également être stationnaire ¹⁶ : on se contente de travailler sur l'approximation AR d'ordre fini pour conclure sur la présence ou non d'une racine unitaire dans le polynôme AR du ARMA(p,q) initial.

3.2.2 à la recherche d'une autre racine unitaire

On va supposer que l'ordre p de la composante autorégressive d'un AR(p) ou d'un ARMA(p,q) est supérieur ou égal à deux. Il possède donc au moins deux racines. Supposons aussi que les tests précédents aient permis de conclure à l'existence d'une racine unitaire, *i.e.* la variable considérée est intégrée d'ordre au moins égal à 1. Il s'agit alors de savoir s'il existe d'autres racines unitaires. Soit l'équation initiale :

$$\begin{aligned}\phi(L)x_t &= u_t, \text{ ou} \\ \phi(L)x_t &= \theta(L)u_t\end{aligned}$$

avec $\phi(L)$ de degré $p \geq 2$ et $\phi(1) = 0$: on a déjà admis, par exemple au moyen des tests ADF, l'existence d'une racine unitaire. On a donc les écritures équivalentes :

$$\begin{aligned}\phi^*(L)\Delta x_t &= u_t, \text{ ou} \\ \phi^*(L)\Delta x_t &= \theta(L)u_t\end{aligned}$$

dans lesquelles $\phi^*(L)$ est un polynôme de degré $p - 1$. On veut tester l'existence d'une racine unitaire pour ce polynôme. La solution est de poser que la variable $y_t = \Delta x_t$ est la nouvelle série d'intérêt. Il suffit donc de faire passer à y_t les tests préalablement réalisés sur x_t . La seule différence peut concerner l'ordre du polynôme autorégressif et donc le nombre d'augmentations du test ADF : partant d'un autorégressif de longueur p , on avait $p - 1$ termes de ce type avec x , on ne devrait en avoir plus que $p - 2$ avec y_t . Deux conclusions sont alors possibles : soit on rejette l'existence d'un trend stochastique dans y , ce qui signifie que $y \sim I(0)$ ou encore que $x \sim I(1)$, soit on ne rejette pas, auquel cas x est intégrée au moins d'ordre 2, et il faut alors tester la présence d'une troisième racine dans la série de travail qui sera alors constituée des différences premières de y , *i.e.* des différences secondes de x . Sachez qu'en pratique, vous trouverez des I(0), plus souvent des I(1), rarement des I(2), et jamais des I(3).

3.2.3 statistiques similaires

Lorsque l'on fait référence au DGP de la variable d'une part et au modèle qui est ajusté d'autre part, une certaine confusion peut apparaître. Peut être que le plus simple est alors de raisonner en termes de *statistiques similaires*. Par définition celles-ci ont un risque qui correspond au risque nominal lorsqu'elles sont utilisées pour le test d'une même hypothèse nulle, même en présence de termes de nuisances qui peuvent être différents dans les modèles ajustés.

Dans le cas qui nous intéresse ici, le test est toujours celui de la nullité du coefficient du niveau retardé dans les différentes régressions vues précédemment, soit $H_0 : b = 0$ avec les notations ci-dessus. Des statistiques seront similaires si, en travaillant avec un seuil de risque α , par exemple égal à 5%, elles conduisent toutes à une probabilité de rejet de 5% lorsque H_0 est vraie et, plus généralement, à une probabilité de rejet égale à α . Naturellement les distributions de statistiques similaires peuvent différer.

La table 1 résume de ce point de vue les caractéristiques des tests DF sachant que ces caractéristiques sont identiques pour les tests ADF. Dans tous les cas il est important de se souvenir que ces tests de Dickey-Fuller sont des tests portant sur la présence de racine unitaire et absolument pas des tests visant à sélectionner une régression plutôt qu'une autre.

15. C'est du moins ce qui ressort des simulations réalisés dans un travail de R. Scott Hacker, "The Effectiveness of Information Criteria in Determining Unit Root and Trend Status", *CESIS, Working Paper 213*, février 2010, et dans celui de R. Scott Hacker & Abdunasser

Data Generating Process	Régressions donnant des tests similaires	statistiques similaires
$x_0 = 0$ et $x_t = x_{t-1} + u_t$	$\Delta x_t = bx_{t-1} + u_t$ $\Delta x_t = bx_{t-1} + \phi_0 + u_t$ $\Delta x_t = bx_{t-1} + \phi_0 + \phi_2 t + u_t$	τ τ_μ τ_t
x_0 quelconque et $x_t = x_{t-1} + u_t$	$\Delta x_t = bx_{t-1} + \phi_0 + u_t$ $\Delta x_t = bx_{t-1} + \phi_0 + \phi_2 t + u_t$	τ_μ τ_t
x_0 quelconque, $c \neq 0$ et $x_t = c + x_{t-1} + u_t$	$\Delta x_t = bx_{t-1} + \phi_0 + \phi_2 t + u_t$	τ_t

TABLE 1 – Statistiques similaires pour $H_0 : b = 0$ dans le cas des tests de Dickey-Fuller

3.3 le test de Phillips-Perron

Afin que les tests τ , τ_μ , τ_t possèdent leur distribution attendue respective, une hypothèse nécessaire est que la série des résidus u_t soit un processus en bruit blanc, *i.e.* composée de variables homoscédastiques orthogonales entre elles. Dans la littérature, il semble que l'accent soit surtout mis sur le respect de l'orthogonalité. Et nous venons de voir que dans le test ADF les augmentations correspondent à une correction paramétrique visant à l'obtention de résidus respectant cette dernière condition.

À la différence de Dickey-Fuller, Phillips et Perron¹⁷ vont adopter une procédure de correction non paramétrique afin de tenir compte d'une éventuelle corrélation sérielle dans la série u_t , cette correction faisant notamment appel à l'estimation de sa variance de long terme, $\hat{\lambda}^2$. On va par ailleurs retrouver les trois configurations déjà vues avec le test DF, à savoir : une régression sans constante et sans trend, $x_t = \phi_1 x_{t-1} + u_t$; une régression avec constante, $x_t = \phi_0 + \phi_1 x_{t-1} + u_t$; une régression avec constante et trend, $x_t = \phi_0 + \phi_1 x_{t-1} + \phi_2 t + u_t$. Dans ce qui suit, k est le nombre de coefficients estimés dans la régression utilisée, *i.e.* k est égal respectivement à 1, 2 ou 3.

Le test de $H_0 : \phi_1 = 1$ versus $H_1 : |\phi_1| < 1$ est réalisé au moyen de la statistique Z_τ définie par¹⁸ :

$$Z_\tau = \sqrt{\frac{\hat{\gamma}_0}{\hat{\lambda}^2}} \frac{(\hat{\phi}_1 - 1)}{s_{\hat{\phi}_1}} - \frac{1}{2} (\hat{\lambda}^2 - \hat{\gamma}_0) \frac{1}{\hat{\lambda}} \frac{T s_{\hat{\phi}_1}}{s_u}$$

$$\hat{\lambda}_T^2 = \hat{\gamma}_{0,T} + 2 \sum_{j=1}^q \left(1 - \frac{j}{q+1}\right) \hat{\gamma}_{j,T}$$

$$\hat{\gamma}_{j,T} = \frac{1}{T} \sum_{t=j+1}^T \hat{u}_t \hat{u}_{t-j}$$

$$s_u^2 = \frac{1}{n-k} \sum_{t=1}^T \hat{u}_t^2$$

Dans ces expressions, la quantité $\frac{(\hat{\phi}_1 - 1)}{s_{\hat{\phi}_1}}$ est donc simplement la valeur du test de student que fournissent les OLS sur le test de $H_0 : \phi_1 = 1$.

Notez qu'à la différence de la variance usuelle de la série des innovations, s_u^2 , la variance de long terme $\hat{\lambda}_T^2$ prend en compte les autocovariances présentes dans cette série. Les pondérations de ces autocovariances sont données par l'estimateur de Newey-West dans lequel la décroissance avec j des poids appliqués à $\hat{\gamma}_{j,T}$ évite l'obtention d'un estimateur qui pourrait être négatif, ce qui pourrait être le cas avec des poids unitaires. En pratique la valeur de cet estimateur $\hat{\lambda}_T^2$ tend à se stabiliser avec l'augmentation du nombre d'autocovariances prises en compte et on peut aisément étudier la sensibilité de la conclusion du test à la valeur donnée à q .

Enfin, la distribution de la statistique Z_τ de Phillips-Perron est la même que celle de la statistique de Dickey-Fuller obtenue dans chacune des 3 configurations possibles soit¹⁹ τ , τ_μ , et τ_t .

Hatemi-J, "Model Selection in Time Series Analysis : Using Information Criteria as an Alternative to Hypothesis Testing", *arXiv preprint, arXiv :1805.08991*, 2018.

16. Notez cependant qu'à distance finie l'équivalence des propriétés du modèle vrai et de son approximation n'est pas vérifiée.

17. Phillips, P.C.B. and P. Perron, Testing for Unit Roots in Time Series Regression, *Biometrika*, 75, 335-346, 1988.

18. Comme pour le test de Dickey-Fuller, nous ne donnons que la statistique de PP construite sur la statistique de student correspondant à l'hypothèse $H_0 : \phi_1 = 1$. Il existe aussi une autre statistique construite sur la quantité $(\hat{\phi}_1 - 1)$.

19. Retenez cependant qu'avec DF et ADF on régresse une variation sur un niveau décalé et d'éventuelles augmentations constituées de différences retardées, chez Phillips-Perron les régressions s'effectuent simplement sur les niveaux de la variable d'intérêt.