

Économétrie des séries temporelles non-stationnaires

Chapitre 1: Les séries temporelles non-stationnaires

Gilles de Truchis

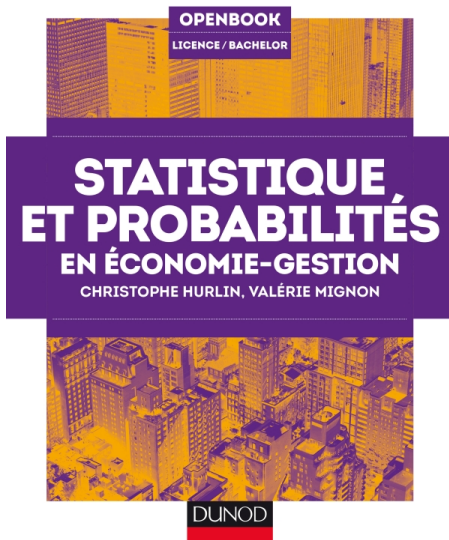
Master 2 EIPMC

Les chapitres du cours

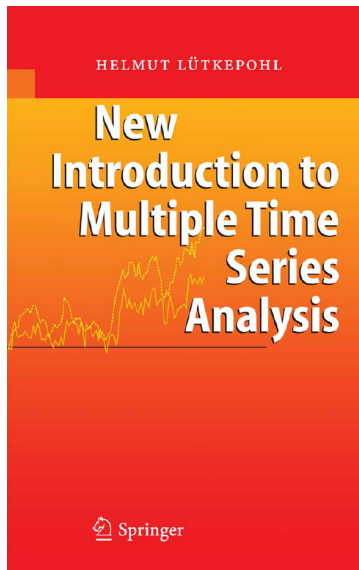
- 1 Faits stylisés et rappels
- 2 Non-stationnarité
- 3 Intégration fractionnaire
- 4 Rappels de Probabilités

- 5 Théorie limite standard
- 6 Théorie limite non-standard
- 7 Les régressions factices
- 8 Conclusion

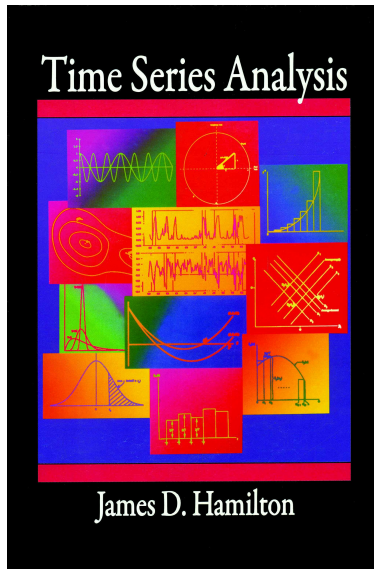
Références



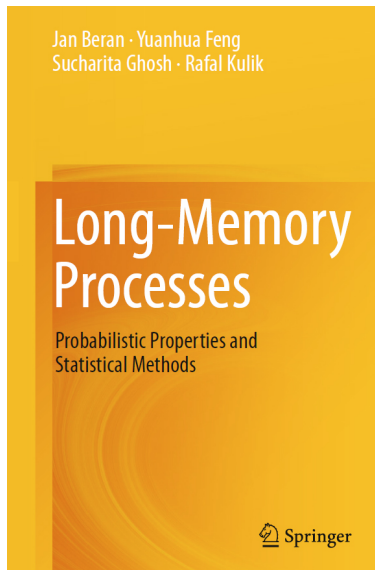
Références



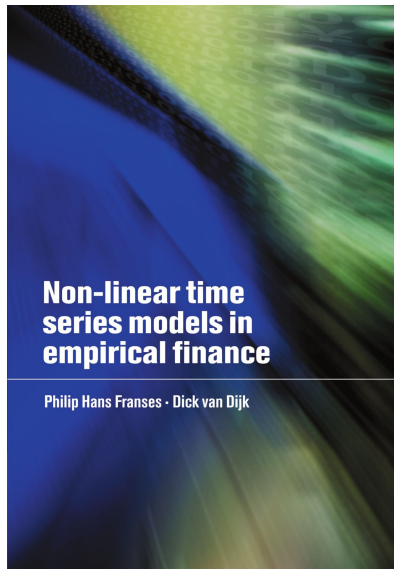
Références



Références



Références



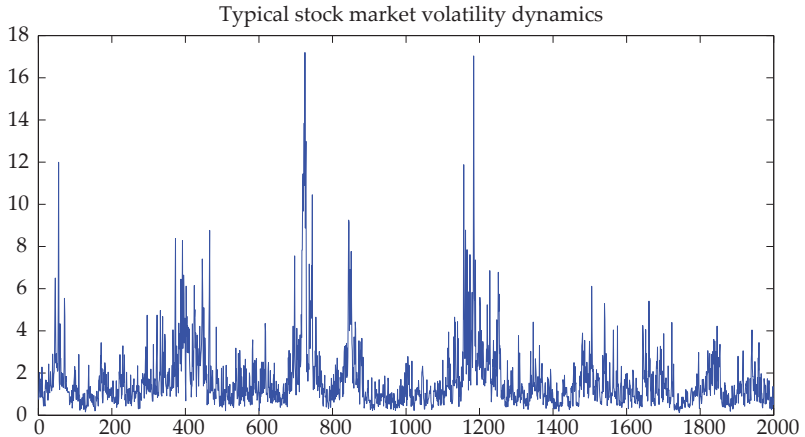
- | | | | |
|---|---------------------------|---|-----------------------------|
| 1 | Faits stylisés et rappels | 5 | Théorie limite standard |
| 2 | Non-stationnarité | 6 | Théorie limite non-standard |
| 3 | Intégration fractionnaire | 7 | Les régressions factices |
| 4 | Rappels de Probabilités | 8 | Conclusion |

- Question centrale : la non-stationnarité des séries temporelles
 - De nombreuses de séries économiques semblent non-stationnaires
 - La non-stationnarité est complexe à définir car protéiforme
 - La non-stationnarité engendre des complications importantes dans la théorie limite des estimateurs
- Quelques exemples :
 - Inflation : Non-stationnarité globale/ locale ? Racine unitaire ?
 - Volatilité financière : Stationnarité ? Non-stationnarité ?

- Difficile de déterminer visuellement si l'inflation est non-stationnaire
 - La fonction d'autocorrélation de l'inflation indique une très forte persistance
 - Pour autant, la série semble stable autour d'une moyenne de long terme

- En fait on verra plus tard qu'avec un processus à mémoire longue ou un AR non-causal on peut approximer ces dynamiques

Variables financières



- Difficile de déterminer visuellement si la volatilité est non-stationnaire
 - La série est animée par de nombreux événements extrêmes
 - Pour autant, la séries semble stable autour d'une moyenne de long terme

- En fait on verra plus tard qu'avec un processus à mémoire longue ou un AR non-causal on peut approximer ces dynamiques

- _____

- C

100% 90% 80% 70% 60% 50% 40% 30% 20% 10% 0%

Définitions

Definition (1)

Un **processus stochastique**, noté $\{X_t(\omega), \omega \in \Omega, t \in \mathbb{R}\}$, est une séquence ordonnée de variables aléatoires définies sur un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$ avec

- Ω l'ensemble **non-dénombrable** des éventualités
- \mathcal{F} un σ -algèbre représentant les événements
- \mathbb{P} une mesure de probabilité telle que $\mathbb{P}(A)$ donne la probabilité de l'événement A

\Rightarrow Par la suite, $\{X_t(\omega), \omega \in \Omega, t \in \mathbb{R}\}$ sera noté $\{X_t\}_{t \in \mathbb{R}}$ ou X_t

Si $\{X_t\}_{t \in \mathbb{R}}$, le processus stochastique est en temps continu, si $\{X_t\}_{t \in \mathbb{Z}}$, il est discret

Definition (2)

Une **série temporelle** notée $\{x_t\}_{t \in \mathbb{T}}$ ou x_t est un segment des réalisations d'un processus stochastique $\{X_t\}_{t \in \mathbb{Z}}$ avec $\mathbb{Z} \supseteq \mathbb{T}$

Definition (3)

Une *série temporelle infinie* notée $\{x_t\}_{t=-\infty}^{\infty}$ est un segment infini des réalisations d'un processus stochastique $\{X_t\}_{t \in \mathbb{Z}}$

le symbol \supseteq dénote un sur-ensemble

Définitions

Definition (4)

*Les **moments inconditionnels ordinaires et centrés** de X_t peuvent s'exprimer comme l'espérance de $h(X_t)$, une fonction continue de X_t*

$$\mathbb{E}(h(X_t)) = \int h(X_t)f(X_t)dX_t$$

avec $f(X_t)$ la fonction de densité inconditionnelle de X_t

- Pour calculer l'espérance de X_t on choisira $h(X_t) = X_t$

$$\mathbb{E}(X_t) = \mu_t$$

- Pour calculer la variance de X_t on choisira $h(X_t) = (X_t - \mathbb{E}(X_t))^2$

$$\mathbb{V}(X_t) = \sigma_t^2$$

Définitions

Definition (5)

La fonction **d'autocovariance** de X_t s'obtient à partir de la densité jointe de $(X_t, X_{t-1}, \dots, X_{t-h})$ et se note

$$\begin{aligned}\gamma(h) &= \text{Cov}(X_t, X_{t-h}) \\ &= \mathbb{E}\left((X_t - \mu_t)(X_{t-h} - \mu_{t-h})\right) \\ &= \int \dots \int (X_t - \mu_t)(X_{t-h} - \mu_{t-h}) f(X_t, \dots, X_{t-h}) dX_t \dots dX_{t-h}\end{aligned}$$

avec $f(X_t, \dots, X_{t-h})$ la fonction de densité inconditionnelle de X_t

Stationnarité au second ordre

- Soit $\{X_t\}_{t \in \mathbb{Z}}$ noté X_t une séquence de variable aléatoire

Definition (6)

X_t est *stationnaire au second ordre* si

- $\forall t \in \mathbb{Z}, \mathbb{E}(X_t) = \mu < \infty$
 - $\forall t, h \in \mathbb{Z}, \text{Cov}(X_t, X_{t+h}) = \gamma(h) < \infty$
- $\Rightarrow \forall t \in \mathbb{Z}, \mathbb{V}(X_t) = \sigma^2 < \infty$ car $\text{Cov}(X_t, X_{t+h}) = \mathbb{V}(X_t)$ pour $h = 0$

- On cherche ici à résumer la stabilité en loi du processus X_t uniquement à travers ses deux premiers moments
 - Pertinent dans le cas Gaussien mais réducteur en générale
 - Cette définition **faible** de la stationnarité est plutôt simple à tester

Stationnarité stricte

- Soit X_t une séquence de variable aléatoire

Definition (7)

X_t est **stationnaire strictement** si la distribution jointe de X_t et X_{t+h} , $\forall t, h$ ne dépend pas de t et uniquement de h

$$f(X_t, \dots, X_{t-h}) = f(X_\tau, \dots, X_{\tau-h})$$

avec $t \neq \tau$

- La distribution jointe du processus X_t doit donc être **invariante par translation dans le temps**
 - Pertinent dans le cas non-Gaussien et Gaussien
 - Cette définition **forte** de la stationnarité est difficile à tester

Ergodicité

- Soit $\{X_t\}_{t=0}^n$ un processus stationnaire au second ordre

Definition (8)

X_t , stationnaire au second ordre est **ergodique** si

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{h=1}^n \gamma(h) = 0$$

avec $\gamma(h)$ sa fonction d'autocovariance

- Cette définition implique que la mémoire du processus est finie
 - une mémoire finie permet d'accumuler de l'information pour l'inférence statistique
 - l'ergodicité décrit une forme faible **d'indépendance asymptotique**
 - **l'absolu sommabilité** de $\gamma(h)$ est une condition suffisante pour

d'indépendance asymptotique : weakly mixing implies ergodicity

Théorème (décomposition) de Wold

Theorem (1 : Décomposition de Wold)

Soit $\{X_t\}_{t=0}^n$ un processus stationnaire au second ordre. On peut montrer que X_t peut toujours se décomposer en une somme pondérée des innovations de X_t et une composante déterministe μ_t

$$X_t = \mu_t + \sum_{j=0}^{\infty} a_j \varepsilon_{t-j}$$

avec $\sum_{j=0}^{\infty} a_j^2 < \infty$ et $\varepsilon_t \sim i.i.d.(0, \sigma_{\varepsilon}^2 < \infty)$

- La convergence en moyenne quadratique des a_j est importante car

$$\mathbb{E}(X_t^2) = \sum_{j=0}^{\infty} a_j^2 \sigma_{\varepsilon}^2 < \infty$$

si pour $m > n$, $\mathbb{E}(\sum_{j=0}^m a_j - \sum_{j=0}^n a_j)^2 < c \Rightarrow \sum_{j=0}^{\infty} a_j^2 < \infty$

- La vitesse de décroissance des coefficients a_j détermine également l'allure de l'autocovariance de x_t car

$$\begin{aligned}\gamma(h) &= \mathbb{E}\left((X_t - \mu_t)(X_{t-h} - \mu_t)\right) \\ &= \mathbb{E}\left(\sum_{m=0}^{\infty} a_m \varepsilon_{t-m} \sum_{s=0}^{\infty} a_{s+h} \varepsilon_{t-s+h}\right) \\ &= \sum_{m,s=0}^{\infty} a_m a_{s+h} \gamma_{\varepsilon}(m-s+h) \\ &= \sigma_{\varepsilon}^2 \sum_{m,s=0}^{\infty} a_m a_{s+h}\end{aligned}$$

ε_t étant un bruit blanc, γ_ε est nul en tout point excepté en 0 où $\gamma_\varepsilon(0) = \sigma_\varepsilon^2$

- | | | | |
|---|---------------------------|---|-----------------------------|
| 1 | Faits stylisés et rappels | 5 | Théorie limite standard |
| 2 | Non-stationnarité | 6 | Théorie limite non-standard |
| 3 | Intégration fractionnaire | 7 | Les régressions factices |
| 4 | Rappels de Probabilités | 8 | Conclusion |

Tendance déterministe

- Il existe de nombreuses façons de violer les hypothèses d'ergodicité et de stationnarité
- Il existe donc tout autant de type de processus non-stationnaires
- Le plus simple d'entre eux : la **tendance déterministe linéaire**
 - $X_t = \mu + \delta t + \varepsilon_t$ avec $\varepsilon_t \sim i.i.d.$ ($0, \sigma_\varepsilon^2 < \infty$)
 - on voit immédiatement que $\mathbb{V}(X_t) = \sigma_\varepsilon^2$ mais $\mathbb{E}(X_t) = \mu + \delta t$
 \Rightarrow l'espérance dépendant du temps, X_t est non-stationnaire
- Ce raisonnement tient pour des fonctions **nonlinéaires** du temps

$$X_t = \mu + \delta(t) + \varepsilon_t$$

avec e.g. $\delta(\cdot)$ une fonction polynomiale

Tendance stochastique et racine unitaire

■ Soit un AR(1) : $X_t = \rho X_{t-1} + \varepsilon_t$, $\varepsilon_t \sim i.i.d. (0, \sigma_\varepsilon^2 < \infty)$

■ Si $\rho = 1$ on constate que $X_t = X_0 + \sum_{j=0}^{t-1} \varepsilon_{t-j}$ et donc

$$\text{Cov}(X_t, X_{t-j}) = (t-j)\sigma_\varepsilon^2 \text{ et } \mathbb{V}(X_t) = \mathbb{V}\left(\sum_{j=0}^{t-1} \varepsilon_{t-j}\right) = t\sigma_\varepsilon^2$$

\Rightarrow La variance de X_t dépend de $t \Rightarrow X_t$ est **non-stationnaire**

$\Rightarrow \rho = 1$ place une solution du polynome de retard sur le cercle unité
donc X_t est un processus **racine unitaire**

$\Rightarrow X_t$ est une **marche aléatoire** d'espérance conditionnelle

$$\mathbb{E}(X_t | X_{t-1}, X_{t-2}, \dots, X_0) = X_{t-1} = \sum_{j=1}^{t-1} \varepsilon_{t-j} = \text{stochastic trend}$$

$\Rightarrow X_t$ est une **martingale de tendance stochastique** $\sum_j \varepsilon_{t-j}$

Solution du polynôme caractéristique $1 - \rho z = 0$

$$z = 1/\rho \Rightarrow \text{ si } \rho = 1 \text{ on a } z = 1$$

Si $|\rho| < 1 \Rightarrow z = |1/\rho| > 1$ la racine est en dehors du cercle unité et le

processus est stationnaire. Si $|\rho| > 1$ la racine est dans le cercle unitaire et le processus admet une solution stationnaire mais non-causale (en temps inversé). Une martingale est un processus stochastique X_t dont l'espérance

en t , sachant l'information passée I_s au temps s , est égale à X_s

$$\mathbb{E}(X_t|I_s) = X_s$$

On parle de sous-martingale si

$$\mathbb{E}(X_t|I_s) \geq X_s$$

On parle de sur-martingale si

$$\mathbb{E}(X_t|I_s) \leq X_s$$

Tendance stochastique et différenciation

- Soit un AR(1) : $X_t = \mu + \rho X_{t-1} + \varepsilon_t$, $\varepsilon_t \sim i.i.d. (0, \sigma_\varepsilon^2 < \infty)$
- Si $\rho = 1$ et $X_0 = 0$, la représentation $MA(\infty)$ nous donne

$$X_t = \rho^t X_0 + \mu \sum_{j=0}^t \rho^j + \sum_{j=0}^{t-1} \rho^j \varepsilon_{t-j} = \mu t + \sum_{j=0}^t \varepsilon_{t-j}$$

- On constate alors que

$$\mathbb{E}(X_t) = \mu t$$

- La **variance** et l'**espérance** ne sont pas indépendantes de t

$\Rightarrow X_t$ est un processus **explosif** de type **marche aléatoire** et de **dérive μ**

- X_t est **stationnaire** en première différence car

$$\Delta X_t = (1 - L)X_t = X_t - X_{t-1} = \mu + \varepsilon_t$$

- Les processus de type tendance **déterministe** ou **stochastique** sont **globalement non-stationnaires**
 - le processus viole les conditions de stationnarité
 - les paramètres du processus sont invariants
- ⇒ la non-stationnarité existe pour toute évolution du processus

- sonst
- re)
- trie

Stationnarité locale

- Les processus dont les paramètres évoluent dans le temps sont possiblement **localement stationnaires**
 - Par exemple, un modèle à coefficients aléatoires dépendant du temps
 - est non-linéaire et **globalement non-stationnaire**
 - peut s'approximer localement par des processus stationnaires
- ⇒ concept de **stationnarité locale** (dépasse le niveau M2)
- Références : Rao (2006), Dahlhaus et Rao (2006)

Plan

- | | | | |
|----------|---------------------------|----------|-----------------------------|
| 1 | Faits stylisés et rappels | 5 | Théorie limite standard |
| 2 | Non-stationnarité | 6 | Théorie limite non-standard |
| 3 | Intégration fractionnaire | 7 | Les régressions factices |
| 4 | Rappels de Probabilités | 8 | Conclusion |

ARMA stationnaire

- Les modèles dynamiques les plus répandus sont les ARMA
- Supposons $X_t \sim ARMA(p, q)$

$$\begin{aligned}
 X_t &= \sum_{j=1}^p \phi_j X_{t-j} + \sum_{j=1}^q \theta_j \varepsilon_{t-j} + \varepsilon_t, \quad \varepsilon_t \sim i.i.d.(0, \sigma_\varepsilon^2 < \infty) \\
 &= \frac{1 - \theta_1 L + \dots + \theta_q L^q}{1 - \phi_1 L + \dots + \phi_p L^p} \varepsilon_t = \frac{\theta(L)}{\phi(L)} \varepsilon_t = \psi(L) \varepsilon_t = \sum_{j=1}^{\infty} a_j \varepsilon_{t-j}
 \end{aligned}$$

dont les racines du polynôme

$$\phi(L) = (1 - \phi_1 L + \dots + \phi_p L^p)$$

ne sont pas situées sur le cercle unité et avec L l'opérateur retard

$\Rightarrow X_t$ est **stationnaire** (voir cours de Pauline Gandré en M1)

Ergodicité des ARMA

- A partir de la forme $MA(\infty)$ de X_t on obtient

$$\gamma(h) = \sigma_\varepsilon^2 \sum_{j=0}^{\infty} a_j a_{j+h}$$

- En valeur absolue et par l'inégalité triangulaire on obtient

$$|\gamma(h)| = \sigma_\varepsilon^2 \left| \sum_{j=0}^{\infty} a_j a_{j+h} \right| \leq \sigma_\varepsilon^2 \sum_{j=0}^{\infty} |a_j a_{j+h}|$$

- On constate alors que les ARMA sont **ergodiques** car

$$\sum_{h=0}^{\infty} |\gamma(h)| \leq \sigma_\varepsilon^2 \sum_{h=0}^{\infty} \sum_{j=0}^{\infty} |a_j a_{j+h}| \quad (1)$$

$$= \sigma_\varepsilon^2 \sum_{j=0}^{\infty} |a_j| \sum_{h=0}^{\infty} |a_{j+h}| < \infty, \quad (2)$$

$\sum_{i=0}^{\infty} |a_j| < \infty$ pour un ARMA et la convolution de deux séries absolument sommables est absolument sommable

Integration

Definition (9)

Un processus stationnaire d'autocovariance $\gamma(h)$ absolument sommable est dit **faiblement dépendant**

Definition (10)

Si un processus est **faiblement dépendant** après δ différenciations, il est dit **intégré d'ordre δ** ou $I(\delta)$

■ Exemple : si $X_t = \mu + X_{t-1} + \varepsilon_t$ avec $\varepsilon_t \sim i.i.d.(0, \sigma_\varepsilon^2 < \infty)$

$\Rightarrow \Delta X_t = \mu + \varepsilon_t$ est **faiblement dépendant**

$\Rightarrow X_t$ est intégré d'ordre 1 également noté $X_t \sim I(1)$

$\Rightarrow \Delta X_t$ est intégré d'ordre nul également noté $\Delta X_t \sim I(0)$

Intégration fractionnaire

- Soit un processus $(1 - L)^\delta X_t = \varepsilon_t$ avec $\delta > -1/2$
 - X_t est un **bruit blanc fractionnaire** si $\varepsilon_t \sim WN(0, \sigma_\varepsilon^2 < \infty)$
- Une représentation $MA(\infty)$ nous donne

$$X_t = (1 - L)^{-\delta} \varepsilon_t = \sum_{j=0}^{\infty} \frac{\Gamma(j + \delta)}{\Gamma(j + 1)\Gamma(\delta)} L^j \varepsilon_t = \sum_{j=0}^{\infty} \alpha_j(-\delta) \varepsilon_{t-j}$$

- Le **polynôme de différenciation fractionnaire** $(1 - L)^{-\delta}$

$$(1 - L)^{-\delta} = 1 + \delta L + \frac{\delta(1 + \delta)}{2!} L^2 + \frac{\delta(1 + \delta)(2 + \delta)}{3!} L^3 + \dots$$

- chaque numérateur est un **symbole de Pochhammer**

$$(\delta)_n = \delta(\delta + 1)(\delta + 2) \dots (\delta + n - 1) = \Gamma(n + \delta)\Gamma(\delta)^{-1}$$

- chaque dénominateur est une **factorielle**, or $n! = \Gamma(n + 1)$

- le j -ième terme de la suite est donc $(\delta)_j \times \Gamma(j + 1)^{-1} L^j$

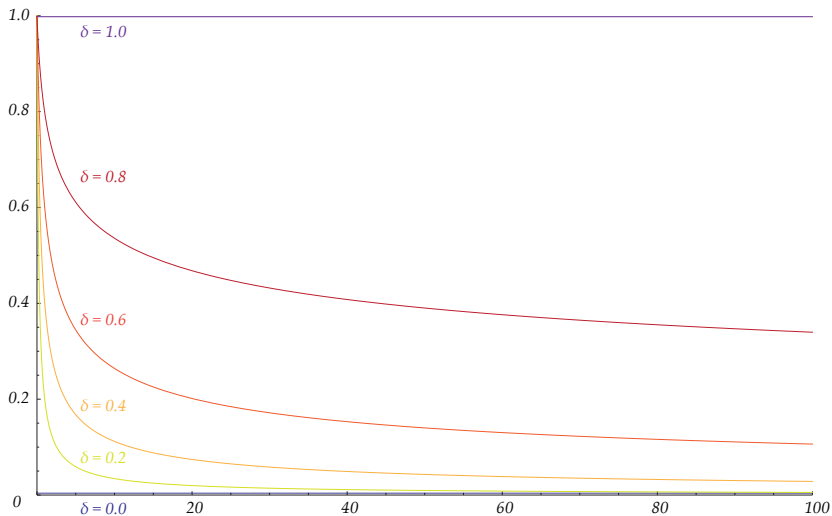
Rappelons que la fonction Gamma est telle que

$$\Gamma(z) = \int_0^{+\infty} t^{z-1} e^{-t} dt$$

et que $\Gamma(z+1) = z\Gamma(z)$ et $\Gamma(1) = 1$

Cette fonction étend la fonction factorielle aux réels et aux complexes

Filtre fractionnaire



Analyse du filtre fractionnaire

- Une analyse à la limite nous apprend que

$$\alpha_j(-\delta) \sim \frac{1}{\Gamma(\delta)} j^{-1+\delta} (1 + O(j^{-1})), \quad j \rightarrow \infty$$

- Granger et Joyeux (1980) soulignent alors que

- si $\delta = 1/2$ on a $\sum_{j=1}^{\infty} \alpha_j(-\delta)^2 = \infty$

\Rightarrow en effet $\alpha_j(-\delta)^2 \approx j^{-1}$ décrit une **série Harmonique** (divergente)

- si $\delta > 1/2$ on a $\sum_{j=1}^{\infty} \alpha_j(-\delta)^2 = \infty$

$\Rightarrow j^{2(\delta-1)=p>-1}$ est une p -série (hyper-Harmonique) divergente

- si $\delta < 1/2$ on a $\sum_{j=1}^{\infty} \alpha_j(-\delta)^2 < \infty$

$\Rightarrow j^{2(\delta-1)=p<-1}$ est une p -série (hyper-Harmonique) convergente

Notation de Landau :

$f(n) = O(g(n))$ veut dire que $f(n)$ est bornée (dominée) par $g(n)$ asymptotiquement ($n \rightarrow \infty$)

$f(n) = o(g(n))$ veut dire que $f(n)$ est négligeable devant $g(n)$ asymptotiquement ($n \rightarrow \infty$)

Bruit blanc fractionnaire et non-stationnarité

- Une représentation $MA(\infty)$ alternative de X_t est alors

$$X_t = \sum_{j=0}^{\infty} \alpha_j(-\delta) \varepsilon_{t-j} \approx \frac{1}{\Gamma(\delta)} \left(\sum_{j=1}^{\infty} j^{-1+\delta} \varepsilon_{t-j} + \varepsilon_t \right)$$

- La variance apparaît alors sous une forme simple

$$\mathbb{V}(X_t) = \sigma_{\varepsilon}^2 \left(\frac{1}{\Gamma^2(\delta)} + \sum_{j=1}^{\infty} \alpha_j(-\delta)^2 \right) = \frac{1}{\Gamma^2(\delta)} \sigma_{\varepsilon}^2 \left(1 + \sum_{j=1}^{\infty} j^{2(\delta-1)} \right)$$

- $\mathbb{V}(X_t)$ sera finie si $\delta < 1/2 \Rightarrow X_t$ **stationnaire**
- $\mathbb{V}(X_t)$ sera infinie si $\delta \geq 1/2 \Rightarrow X_t$ **non-stationnaire**
- Il existe une formulation non-asymptotique de la variance

$$\mathbb{V}(X_t) = \sigma_{\varepsilon}^2 \frac{\Gamma(1-2\delta)}{\Gamma^2(1-\delta)} = \sigma_{\varepsilon}^2 v_{\delta}$$

où $v_{\delta} = 1$ si $\delta = 0$ et $v_{\delta} = \infty$ si $\delta \geq 1/2$

Bruit blanc fractionnaire et autocovariance

- Les calculs d'autocovariance obtenus par Hosking (1981) donnent

$$\begin{aligned}\gamma(h) &= \mathbb{E}\left((X_t X_{t-h})\right) \\ &= \sigma_\varepsilon^2 \frac{\Gamma(h+\delta)\Gamma(1-2\delta)}{\Gamma(h+1-\delta)\Gamma(1-\delta)\Gamma(\delta)} \\ &= \frac{(\sigma_\varepsilon^2/2\pi) \sin(\pi\delta)\Gamma(h+\delta)\Gamma(1-2\delta)}{\Gamma(h+1-\delta)}\end{aligned}$$

avec $\delta \in (-1/2, 1/2)$

- Approximation asymptotique de Lieberman et Phillips (2008)

$$\gamma(h) \sim \frac{\sigma_\varepsilon^2}{\pi} \frac{\Gamma(1-2\delta) \sin(\pi\delta)}{h^{1-2\delta}} + O(h^{2\delta-3}), \quad h \rightarrow \infty$$

avec $\delta \in (-1/2, 0) \cup (0, 1/2)$

Bruit blanc fractionnaire et sommabilité

- Via Lieberman et Phillips (2008) on constate que

$$\gamma(h) \sim Ch^{2\delta-1}, \quad C > 0$$

- Or pour $\delta \in (0, 1/2)$ on a $2\delta - 1 \in (-1, 0)$ et donc

$$\sum_{j=0}^{\infty} |\gamma(j)| = \infty \text{ bien que } \mathbb{V}(X_t) < \infty$$

- En revanche, pour $\delta \in (-1/2, 0)$, $2\delta - 1 \in (-2, -1)$ et donc

$$\sum_{j=0}^{\infty} |\gamma(j)| < \infty$$

- Si $\delta \in (-1/2, 0)$, $\gamma(j)$ est absolument sommable

$\Rightarrow X_t$ est dit **faiblement dépendant** et à **mémoire courte**

- Si $\delta \in (0, 1/2)$, X_t , $\gamma(j)$ n'est pas absolument sommable

$\Rightarrow X_t$ est dit **fortement dépendant** et à **mémoire longue**

Bruit blanc fractionnaire et mean reversion

Definition (12)

Un processus X_t est **mean-reverting** (revient vers sa moyenne) si sa fonction de réponse impulsionnelle cumulée (CIR) à l'infini tend vers zéro.

- Soit $(1 - L)^\delta X_t = \varepsilon_t$ un bruit blanc fractionnaire avec $\delta > 1/2$
- Considérons à présent la représentation $MA(\infty)$ de $(1 - L)X_t$

$$(1 - L)X_t = (1 - L)^{1-\delta} \varepsilon_t$$

- On a alors l'impact d'un choc unitaire en t sur X en $t + h$

$$CIR_h = \sum_{j=0}^h \alpha_j (1 - \delta) \sim C \sum_{j=1}^h j^{\delta-2}, \quad C > 0$$

dont la convergence est vérifiée pour $\delta < 1$ quand $h \rightarrow \infty$

- On peut généraliser le bruit blanc fractionnaire en ajoutant de la dynamique de court terme

Definition (13)

Un processus X_t est un ARFIMA(p, δ, q) si une fois différencié δ fois il peut s'exprimer comme un processus stationnaire et invertible de type ARMA(p, q)

$$\phi(L)\Delta^\delta X_t = \alpha + \theta(L)\varepsilon_t$$

- Les démonstrations précédentes tiennent pour un ARFIMA bien que les formules se complexifient

ARFIMA

■ A la différence des processus ARIMA(p, δ, q)

$\Rightarrow X_t$ est stationnaire à mémoire longue si $\delta < 1/2$

$\Rightarrow X_t$ est non-stationnaire à mémoire longue si $\delta > 1/2$

$\Rightarrow X_t$ est non-stationnaire mean-reverting si $1/2 \leq \delta < 1$

$\Rightarrow X_t$ est stationnaire à mémoire courte si $\delta \leq 0$

$\Rightarrow X_t$ est dit anti-persistant si $\delta < 0$

$\Rightarrow X_t$ possède une représentation MA(∞) si $\delta > -1/2$

$\Rightarrow X_t$ possède une représentation AR(∞) si $\delta < 1/2$

■ Les deux derniers points sont discutés par Hosking (1981)

Les concepts de convergences

- Soit X_i une fonction de n variables aléatoires $X_n = f(Y_1, \dots, Y_n)$
 - L'étude du comportement de X_n quand $n \rightarrow \infty$ est cruciale
 - $f(\cdot)$ sera souvent un **estimateur**
 - L'étude de ce comportement **limite** repose sur différentes notions de convergence
 - convergence **presque sûre**
 - convergence en **probabilité**
 - convergence en **moyenne quadratique**
 - convergence en **loi**

Convergence presque sûre

- **Implications** : quand $n \rightarrow \infty$, X_n tend de façon certaine vers une constante (i.e. une variable aléatoire dégénérée)

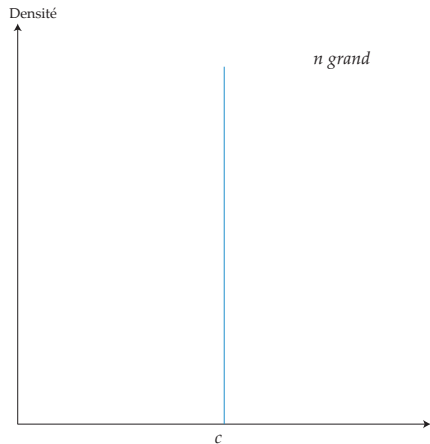
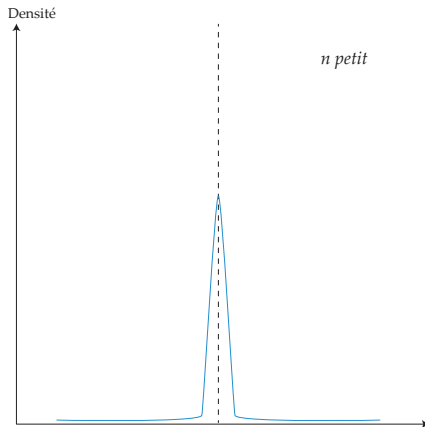
Definition (14)

x_n converge presque sûrement vers une constante c si,

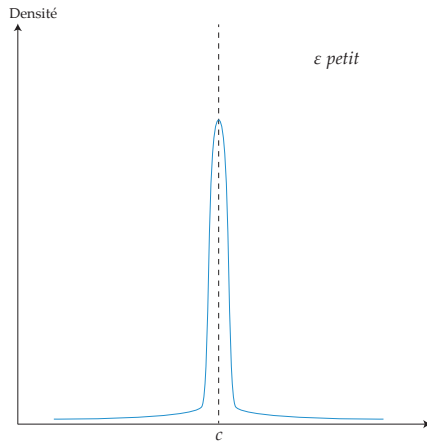
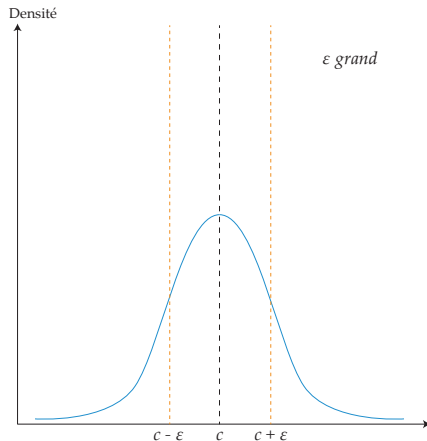
$$\Pr \left(\lim_{n \rightarrow \infty} X_n = c \right) = 1$$

- **Notation mathématique** : $X_n \xrightarrow{a.s.} c$
- **Explications** : Comme X_n tend vers une valeur **constante** de manière **certaine**, sa **distribution asymptotique** est une **masse ponctuelle**

Convergence presque sûre : Illustration



Convergence en probabilité : Illustration



Convergence en loi

- **Implications** : quand $n \rightarrow \infty$, X_n tend vers une autre variable aléatoire et dont la distribution est asymptotiquement équivalente

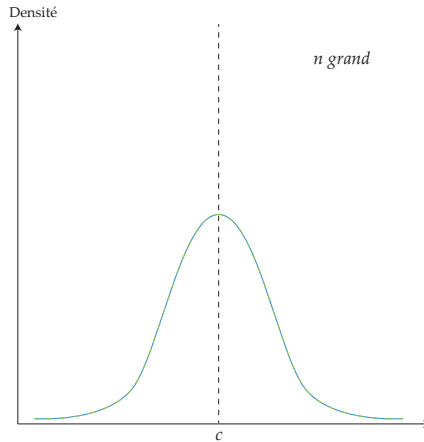
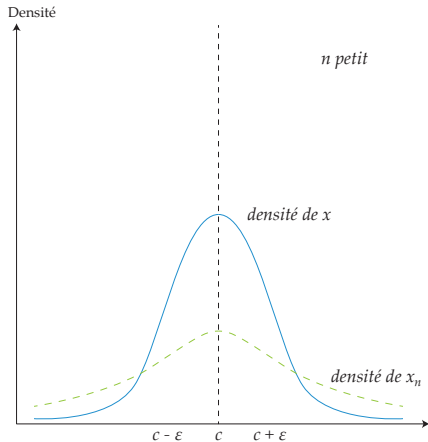
Definition (17)

Soit $F_n(\cdot)$ la fonction de répartition de X_n . X_n converge en loi vers une variable aléatoire X définie sur un support $X(\Omega)$ et ayant pour fonction de répartition $F(\cdot)$ si,

$$\lim_{n \rightarrow \infty} F_n(z) = F(z), \quad \forall z \in X(\Omega)$$

- **Notation mathématique** : $X_n \xrightarrow{d} X$ ou $X_n \xrightarrow{d} \mathcal{L}$
- **Explications** : Asymptotiquement la distribution de X_n est donc identique à celle de X , ce qui implique des fonctions de densité et de répartition identiques : X_n et X sont identiquement distribuées

Convergence en loi : Illustration



Loi faible des grands nombres

- **Implications** : il s'agit d'un théorème portant sur une séquence de variables aléatoires i.i.d.

Theorem (2 : Weak Law of Large Numbers)

Pour une séquence de variables aléatoires i.i.d. , $X_t = X_1, \dots, X_n$, la moyenne empirique de ces variables converge en probabilité vers l'espérance de X_t

$$\bar{X} = \frac{1}{n} \sum_{t=1}^n X_t \xrightarrow{p} \mathbb{E}(X_t)$$

- Par la suite on notera **LLN** la loi des grands nombres

Loi forte des grands nombres

- **Notes :** Il existe des versions fortes de cette loi

$$\text{plim } \bar{X} = \lim_{n \rightarrow \infty} n^{-1} \sum_{t=1}^n \mathbb{E}(X_t) \text{ si } X_t \text{ indépendant } \forall t \quad (3)$$

$$= \mu \text{ si } X_t \sim \text{i.i.d.} \quad (4)$$

Theorem (3 : LLN Kolmogorov)

Pour une séquence de variables aléatoires i.i.d. , $X_t = X_1, \dots, X_n$, si $\mathbb{E}(|X_t|) < \infty$, la moyenne empirique de ces variables converge presque sûrement vers l'espérance de X_t

$$\bar{X} = \frac{1}{n} \sum_{t=1}^n X_t \xrightarrow{a.s.} \mathbb{E}(X_t)$$

Loi forte des grands nombres

- Le théorème LLN de Markov relâche l'hypothèse de distribution identique au coût de conditions sur les moments plus élevés

Theorem (4 : LLN Markov)

Pour une séquence de variables aléatoires indépendamment mais non-identiquement distribuées, $X_t = X_1, \dots, X_n$, avec $\mathbb{E}(X_t) = \mu_t$ et $\mathbb{V}(X_t) = \sigma_t^2$, si

$$\sum_{t=1}^{\infty} \left(\mathbb{E}(|X_t - \mu_t|^{1+m}) / t^{1+m} \right) < \infty$$

pour $m > 0$, alors

$$\bar{X} = \frac{1}{n} \sum_{t=1}^n X_t \xrightarrow{a.s.} \frac{1}{n} \sum_{t=1}^n \mathbb{E}(X_t)$$

Vers un théorème de la limite centrée

- Soit une séquence i.i.d. , $X_t = X_1, \dots, X_n$ de moyenne

$$\bar{X} = n^{-1} \sum_{t=1}^n X_t$$

- D'après la LLN de Kolmogorov, $\bar{X} \xrightarrow{a.s.} \mathbb{E}(X_t)$ si $n \rightarrow \infty$

- la distribution asymptotique de \bar{X} est dégénérée

\Rightarrow Comment construire une statistique inférentielle ?

- La logique consiste à opérer une transformation de \bar{X} tel que

$$\mathbb{T}(\bar{X}) \xrightarrow{d} \mathcal{L}$$

avec \mathcal{L} une distribution non dégénérée

- Cette transformation sera généralement de la forme

$$\mathbb{T}(\bar{X}) = \sqrt{n} (\bar{X} - \mathbb{E}(X_t))$$

Pourquoi une telle transformation ?

- Pour simplifier supposons $\mathbb{E}(X_t) = 0$ et $\mathbb{T}(\bar{X}) = n^\alpha \bar{X}$

- Comme $X_t \sim \text{i.i.d.}$, $\mathbb{E}(X_t) = 0$, $\mathbb{V}(X_t) = \sigma_X^2$ et $\mathbb{Cov}(X_i, X_j) = 0$, $i \neq j$

- On en déduit alors

$$\mathbb{E}(n^\alpha \bar{X}) = n^\alpha \mathbb{E}(\bar{X}) = 0, \quad \mathbb{V}(n^\alpha \bar{X}) = n^{2\alpha-1} \sigma_X^2$$

- Car :

$$\begin{aligned} \mathbb{V}(\bar{X}) &= \mathbb{V}\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n^2} \mathbb{V}\left(\sum_{i=1}^n X_i\right) \\ &= \frac{1}{n^2} \left(\sum_{i=1}^n \mathbb{V}(X_i) + 2 \sum_{i=1}^n \sum_{j=i+1}^{n-1} \mathbb{Cov}(X_i, X_j) \right) \\ &= \frac{1}{n^2} \sum_{i=1}^n \mathbb{V}(X_i) = \frac{1}{n^2} \sum_{i=1}^n \sigma_X^2 = \frac{n \sigma_X^2}{n^2} = \frac{\sigma_X^2}{n} \end{aligned}$$

La vitesse de convergence

- Pour $\alpha \geq 0$ considérons trois cas

- $\alpha > 1/2$ et donc $2\alpha - 1 > 0$

$$\lim_{n \rightarrow \infty} \mathbb{V}(n^\alpha \bar{X}) = \sigma^2 \lim_{n \rightarrow \infty} n^{2\alpha-1} = \infty$$

- $\alpha < 1/2$ et donc $2\alpha - 1 < 0$

$$\lim_{n \rightarrow \infty} \mathbb{V}(n^\alpha \bar{X}) = \sigma^2 \lim_{n \rightarrow \infty} n^{2\alpha-1} = 0$$

- $\alpha = 1/2$ et donc $2\alpha - 1 = 0$

$$\lim_{n \rightarrow \infty} \mathbb{V}(n^\alpha \bar{X}) = \sigma^2 \lim_{n \rightarrow \infty} n^{2\alpha-1} = \sigma^2$$

\Rightarrow La **normalisation** \sqrt{n} préserve la variance de X_t

\Rightarrow On dit que $\bar{X} - \mathbb{E}(X_t)$ converge à la vitesse \sqrt{n}

Le théorème central limite

Theorem (5 : TCL de Lindeberg-Levy)

Soit une séquence i.i.d. , $X_t = X_1, \dots, X_n$ d'espérance $\mathbb{E}(X_t) = m$ et de variance finies $\mathbb{V}(X_t) = \sigma^2$. D'après le théorème central limite de Lindeberg-Levy,

$$\tilde{Z}_n = \sqrt{n} (\bar{X} - m) \xrightarrow{d} \mathcal{N}(0, \sigma^2) \equiv Z_n = \frac{\tilde{Z}_n}{\sigma} \xrightarrow{d} \mathcal{N}(0, 1)$$

avec $n \rightarrow \infty$ et \tilde{Z}_n s'exprimant également comme

$$\sqrt{n} \frac{1}{n} \sum_{t=1}^n (X_t - m) = n^{-1/2} \sum_{t=1}^n (X_t - m) = \sqrt{n} \left(\underbrace{\left(\frac{1}{n} \sum_{t=1}^n X_t \right)}_{\bar{X}} - \underbrace{\frac{1}{n} nm}_m \right)$$

- D'autres théorèmes centraux limites existent
 - TCL de Lyapunov (développé par la suite)
 - TCL pour martingales et processus mélangeants (non présentés)

Le théorème central limite

Theorem (6 : TCL de Lyapunov)

Soit une séquence i. ni. d. , $X_t = X_1, \dots, X_n$ d'espérance $\mathbb{E}(X_t) = \mu_t$ et de variance finies $\mathbb{V}(X_t) = \sigma_t^2$. D'après le théorème central limite de Lyapunov, si

$$\lim_{n \rightarrow \infty} \frac{1}{s_n^{2+m}} \left(\sum_{t=1}^n \mathbb{E}(|X_t - \mu_t|^{2+m}) \right) < \infty, \quad s_n^2 = \sum_{t=1}^n \sigma_t^2$$

pour $m > 0$,

$$Z_n = \frac{\sum_{t=1}^n (X_t - \mu_t)}{\sqrt{\sum_{t=1}^n \sigma_t^2}} \xrightarrow{d} \mathcal{N}(0, 1)$$

La distribution asymptotique

- Soit une séquence i.i.d. , $X_t = X_1, \dots, X_n$ convergeant en loi vers X , une variable ayant pour fonction de répartition $F(\cdot)$

$$X_n \xrightarrow{d} \mathcal{L}_X$$

- $F(\cdot)$ est donc la fonction de répartition de la **distribution asymptotique** de X_n
- Dans le cadre du TCL, supposons

$$\sqrt{n}(X_n - m) \xrightarrow{d} \mathcal{N}(0, \sigma^2)$$

- Peut-on en conclure que $X_n \xrightarrow{d} \mathcal{N}(m, \sigma^2/n)$?

\Rightarrow En fait, on peut uniquement dire que

$$X_n \xrightarrow{a.a.d} \mathcal{N}(m, \sigma^2/n)$$

et que $\mathbb{V}_{asy}(X_n) = \sigma^2/n$ et $\mathbb{E}_{asy}(X_n) = m$

$\xrightarrow{a.a.d}$ veut dire approximativement asymptotiquement distribué

Économétrie des séries temporelles non-stationnaires

- | | | | |
|---|---------------------------|---|-----------------------------|
| 1 | Faits stylisés et rappels | 5 | Théorie limite standard |
| 2 | Non-stationnarité | 6 | Théorie limite non-standard |
| 3 | Intégration fractionnaire | 7 | Les régressions factices |
| 4 | Rappels de Probabilités | 8 | Conclusion |

Rappels sur les OLS

- La non-stationnarité impact la **théorie limite** des estimateurs
- Pour comprendre cela repartons des OLS dans le modèle linéaire

$$Y_t = X_t\beta + \varepsilon_t \quad \varepsilon_t \sim \text{i.i.d. } (0, \sigma_\varepsilon^2)$$

- Supposons $X_t \perp \varepsilon_t$
- L'estimateur OLS est alors donné par

$$\begin{aligned} \hat{\beta} &= \frac{\text{Cov}(Y, X)}{\mathbb{V}(X)} = \left(\frac{1}{n} \sum_{t=1}^n X_t^2 \right)^{-1} \frac{1}{n} \sum_{t=1}^n X_t Y_t \\ &= \left(\frac{1}{n} \sum_{t=1}^n X_t^2 \right)^{-1} \frac{1}{n} \sum_{t=1}^n (\beta X_t + \varepsilon_t) X_t = \beta + \frac{n^{-1} \sum_{t=1}^n X_t \varepsilon_t}{n^{-1} \sum_{t=1}^n X_t^2} \end{aligned}$$

Consistance des OLS

- Analysons la **consistance** de $\hat{\beta}$ si $X_t \sim$ i. ni. d. $(0, \sigma_X^2 < \infty)$

$$\text{plim } \hat{\beta} = \beta + \frac{\text{plim } n^{-1} \sum_{t=1}^n X_t \varepsilon_t}{\text{plim } n^{-1} \sum_{t=1}^n X_t^2}$$

- On observe que $\mathbb{E}(X_t \varepsilon_t) = \mathbb{E}(X_t) \mathbb{E}(\varepsilon_t) = 0$ et $\mathbb{V}(X_t \varepsilon_t) = \sigma_X^2 \sigma_\varepsilon^2$

- $\mathbb{V}(X_t \varepsilon_t) = \mathbb{E}(X_t)^2 \mathbb{E}(\varepsilon_t)^2 - \mathbb{E}(X_t)^2 \mathbb{V}(\varepsilon_t) - \mathbb{E}(\varepsilon_t)^2 \mathbb{V}(X_t) + \mathbb{V}(X_t) \mathbb{V}(\varepsilon_t)$

- dans la LLN de Markov (4), $\sum \sigma_X^2 \sigma_\varepsilon^2 / t^2 < \infty$ pour $m = 1$

$$\Rightarrow n^{-1} \sum_{t=1}^n X_t \varepsilon_t \xrightarrow{p} \mathbb{E}(X_t \varepsilon_t) = 0$$

- On suppose que $\mathbb{E}(X_t^4) < \infty$ existe et par la LLN de Markov,

$$n^{-1} \sum_{t=1}^n X_t^2 \xrightarrow{p} \sigma_X^2$$

- Par application du théorème de **Mann–Wald** (7) on a

$$\text{plim } \hat{\beta} = \beta + \frac{0}{\text{plim } n^{-1} \sum_{t=1}^n X_t^2} = \beta$$

Si X_t non-stochastique et $\varepsilon_t \sim$ i.i.d. alors $(X_t \varepsilon_t)$ est indépendant mais pas identiquement distribué

La démonstration avec X_t stochastique nécessite $\mathbb{V}(X_t) < \infty$ et une application de la LLN de Markov sur X_t^2 et donc que $\mathbb{E}(X_t^4)$ existe et soit fini.

Distribution limite des OLS

- La consistance donne une distribution dégénérée : $\hat{\beta} \xrightarrow{p} \beta$

- Avec un transformation $\mathbb{T}(\hat{\beta}) \xrightarrow{d} \mathcal{L}$ on peut montrer que

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} \mathcal{N}(0, \sigma_{\varepsilon}^2)$$

- Commençons par analyser $n^{-1} \sum_{t=1}^n X_t \varepsilon_t$ sachant $\mathbb{E}(X_t \varepsilon_t) = 0$ et

$$\mathbb{V}(X_t \varepsilon_t) = \sigma_X^2 \sigma_{\varepsilon}^2$$

\Rightarrow L'application du TCL de **Lyapunov** (6) donne

$$\sqrt{n}(n^{-1} \sum_{t=1}^n X_t \varepsilon_t - 0) = n^{-1/2} \sum_{t=1}^n X_t \varepsilon_t \xrightarrow{d} \mathcal{N}(0, \sigma_X^2 \sigma_{\varepsilon}^2)$$

- La consistance a montré que $n^{-1} \sum_{t=1}^n X_t^2 \xrightarrow{p} \sigma_X^2$

\Rightarrow D'après le théorème de **Slutsky** (8) on a alors

$$\sqrt{n}(\hat{\beta} - \beta) = \sqrt{n} \frac{n^{-1} \sum_{t=1}^n X_t \varepsilon_t}{n^{-1} \sum_{t=1}^n X_t^2} \xrightarrow{d} \frac{\mathcal{N}(0, \sigma_X^2 \sigma_{\varepsilon}^2)}{\sigma_X^2} \xrightarrow{d} \mathcal{N}(0, \sigma_{\varepsilon}^2 \sigma_X^{-2})$$

Théorie limite dans le cas d'un AR(1)

■ Soit $X_t = \rho X_{t-1} + \varepsilon_t$, $\varepsilon_t \sim \text{i.i.d. } (0, \sigma_\varepsilon^2)$ et $X_0 = 0$

■ L'estimateur OLS est donné par

$$\hat{\rho} = \rho + \left(\sum_{t=1}^n X_t^2 \right)^{-1} \left(\sum_{t=1}^n X_{t-1} \varepsilon_t \right)$$

■ Sous l'hypothèse que $|\rho| < 1$, X_t est stationnaire et

$$\frac{1}{n} \sum_{t=1}^n X_t^2 \xrightarrow{p} \sigma_X^2 = \mathbb{E} \left(\frac{1}{n} \sum_{t=1}^n X_t^2 \right) = \frac{\sigma_\varepsilon^2}{1 - \rho^2} < \infty$$

\Rightarrow D'après le théorème de **Slutsky** (8) on a alors

$$\sqrt{n}(\hat{\rho} - \rho) \xrightarrow{d} \frac{\mathcal{N}(0, \sigma_\varepsilon^2 \sigma_X^2)}{\sigma_X^2} \stackrel{d}{=} \mathcal{N}(0, 1 - \rho^2) \quad (5)$$

■ Qu'observez-vous si $\rho = 1$?

- Décomposons

$$\mathbb{E}\left(\frac{1}{n} \sum_{t=1}^n X_t^2\right) = \frac{1}{n} \sum_{t=1}^n \mathbb{E}(X_t^2) = \frac{1}{n} \sum_{t=1}^n \mathbb{E}\left((\rho X_{t-1} + \varepsilon_t)^2\right)$$

dont le développement de l'identité remarquable se simplifie (car le terme de covariance $2\rho X_{t-1}\varepsilon_t$ est nul) telle que

$$\sigma_X^2 = \frac{1}{n} \sum_{t=1}^n \mathbb{E}(\rho^2 X_{t-1}^2 + \varepsilon_t^2) = \rho^2 \sigma_X^2 + \sigma_\varepsilon^2 = \frac{\sigma_\varepsilon^2}{1 - \rho^2}$$

- D'après le théorème de **Slutsky** on a

$$\sqrt{n}(\hat{\rho} - \rho) \xrightarrow{d} \mathcal{N}(0, v)$$

avec $\sigma_X^2 = \sigma_\varepsilon^2(1 - \rho^2)^{-1}$ et

$$v = \frac{\sigma_\varepsilon^2 \sigma_X^2}{(\sigma_X^2)^2} = \frac{\sigma_\varepsilon^2 \sigma_\varepsilon^2 (1 - \rho^2)^{-1}}{\sigma_\varepsilon^4 (1 - \rho^2)^{-2}} = 1 - \rho^2$$

- On remarque que si $\rho = 1$, la distribution à une variance nulle... Elle est donc dégénérée et $\sqrt{n}(\hat{\rho} - 1) \xrightarrow{p} 0$. Dans le cas racine unitaire, les OLS sont donc consistants mais cette théorie limite n'est d'aucune aide pour des tests d'hypothèses sur $\hat{\rho}$

- | | | | |
|---|---------------------------|---|-----------------------------|
| 1 | Faits stylisés et rappels | 5 | Théorie limite standard |
| 2 | Non-stationnarité | 6 | Théorie limite non-standard |
| 3 | Intégration fractionnaire | 7 | Les régressions factices |
| 4 | Rappels de Probabilités | 8 | Conclusion |

Limite de la théorie asymptotique standard

■ Soit $X_t = \rho X_{t-1} + \varepsilon_t$, $\varepsilon_t \sim \text{i.i.d. } (0, \sigma_\varepsilon^2)$, $X_0 = 0$ et $\rho = 1$

■ X_t est donc une **marche aléatoire de variance** $\mathbb{V}(X_t) = t\sigma_\varepsilon^2$

■ Impossible alors d'appliquer un TCL comme dans (5) puisque

$$\frac{1}{n} \sum_{t=1}^n X_t^2 \xrightarrow{p} \mathbb{E} \left(\frac{1}{n} \sum_{t=1}^n X_t^2 \right) = \frac{1}{n} \sum_{t=1}^n \mathbb{E}(X_t^2) = \sigma_\varepsilon^2 \frac{\sum_{t=1}^n t}{n} \rightarrow \infty$$

■ Pour autant une **théorie limite non-standard** est possible si

$$\mathbb{T}(\hat{\rho}) = n(\hat{\rho} - \rho) \neq \sqrt{n}(\hat{\rho} - \rho)$$

\Rightarrow Théorie asymptotique applicable dans le cas **non-stationnaire**

■ Ce type de théorie fait intervenir les **processus de Wiener**

Marche aléatoire et TCL

- Soit $X_t = \rho X_{t-1} + \varepsilon_t$, $\varepsilon_t \sim \text{i.i.d. } (0, \sigma_\varepsilon^2)$ et $X_0 = 0$
- On a vu que si $\rho = 1$ alors $X_n = \sum_{t=1}^n \varepsilon_t$
- On considère à présent

$$\mathbb{T}(X_n) = \sqrt{n} \frac{1}{n} X_n \sigma_{\mathbb{T}(X)}^{-1} = \sqrt{n}(\bar{\varepsilon}) \sigma_{\mathbb{T}(X)}^{-1} = n^{-1/2} X_n \sigma_{\mathbb{T}(X)}^{-1}$$

avec

$$\sigma_{\mathbb{T}(X)}^2 = \mathbb{V}(\mathbb{T}(X_n)) = \mathbb{E} \left(n^{-1} \left(\sum_{t=1}^n \varepsilon_t \right)^2 \right) = n^{-1} n \sigma_\varepsilon^2 = \sigma_\varepsilon^2 \quad (6)$$

- D'après le TCL de Lindeberg-Lévy (5), on constate alors que

$$n^{-1/2} X_n \sigma_{\mathbb{T}(X)}^{-1} \xrightarrow{d} \mathcal{N}(0, 1)$$

Les sommes partielles

- Poursuivons ce raisonnement sur une somme partielle de X_n

$$X_n(r) = \sum_{s=1}^{[nr]} \varepsilon_s$$

avec $0 \leq r < 1$ et $[nr]$ le plus grand entier $\leq nr$

- Après transformation de $X_n(r)$ on a donc

$$W_n(r) = n^{-1/2} X_n(r) \sigma_{\mathbb{T}(X)}^{-1} = n^{-1/2} \sum_{s=1}^{[nr]} \varepsilon_s \sigma_{\mathbb{T}(X)}^{-1}$$

... que l'on peut réécrire

$$W_n(r) = \underbrace{\left(n^{-1/2} [nr]^{1/2} \right)}_{\rightarrow r^{1/2}} \times \underbrace{\left([nr]^{-1/2} \sum_{s=1}^{[nr]} \varepsilon_s \sigma_{\mathbb{T}(X)}^{-1} \right)}_{\xrightarrow{d} \mathcal{N}(0,1)}$$

$$\Rightarrow W_n(r) \xrightarrow{d} r^{1/2} \mathcal{N}(0, 1) \xrightarrow{d} \mathcal{N}(0, r)$$

Vers les processus de Wiener

- A présent définissons $r_t = t/n$ de telle sorte que $[nr_t] = t$ et

$$W_n(r) = n^{-1/2} \sum_{s=1}^t \varepsilon_s \sigma_{\mathbb{T}(X)}^{-1}$$

- Si $t/n < r_i < (t+1)/n$ on a également $[nr_i] = t$ et donc

$$W_n(r) = n^{-1/2} \sum_{s=1}^t \varepsilon_s \sigma_{\mathbb{T}(X)}^{-1}$$

- Mais si $r_i = (t+1)/n$ on a

$$W_n(r) = n^{-1/2} \sum_{s=1}^{t+1} \varepsilon_s \sigma_{\mathbb{T}(X)}^{-1}$$

\Rightarrow par définition $r \in [0, 1]$ et

$$W_n(1) = n^{-1/2} \sum_{s=1}^n \varepsilon_s \sigma_{\mathbb{T}(X)}^{-1} \quad (7)$$

Les processus de Wiener sont également connus sous le nom de mouvement Brownien. Ce sont des processus continus, normalement distribués et leur variance dépend du temps $t \in [0, 1]$

Théorème central limite fonctionnel

- $W_n(r)$ est une **fonction aléatoire** (via ε_t) de r

Theorem (9 : Théorème de Donsker)

Soit ε_t une séquence de variables aléatoires telle que $\varepsilon_t \sim$ i.i.d. $(0, \sigma_\varepsilon^2 < \infty)$, alors

$$W_n \xrightarrow{d} W$$

- Le TCL fonctionnel requiert (pas suffisant) la convergence point par point de $W_n(r)$

$$W_n(r) \xrightarrow{d} W(r)$$

Dans ce qui suit, par abus de notation nous parlons de convergence en loi \xrightarrow{d} pour parler d'une notion plus large et complexe, la convergence faible. Mais nous n'attendrons pas ce niveau de détail.

Théorème de Mann-Wald fonctionnel

Theorem (10 : Functional Continuous Mapping Theorem)

Soit $X_n(.) \xrightarrow{d} X(.)$ une fonction aléatoire convergente et $g(.)$ une fonction continue à valeur dans \mathbb{R} en $X(.)$. Alors,

$$X_n(.) \xrightarrow{d} X(.) \Rightarrow g(X_n) \xrightarrow{d} g(X)$$

$$X_n(.) \xrightarrow{p} X(.) \Rightarrow g(X_n) \xrightarrow{p} g(X)$$

$$X_n(.) \xrightarrow{a.s.} X(.) \Rightarrow g(X_n) \xrightarrow{a.s.} g(X)$$

Théorie limite non standard des OLS

Theorem (11)

Soit $X_t = \rho X_{t-1} + \varepsilon_t$, $\varepsilon_t \sim \text{i.i.d. } (0, \sigma_\varepsilon^2)$ et $X_0 = 0$. Si $\rho = 1$, la distribution limite de l'estimateur OLS est donnée par

$$n(\hat{\rho} - \rho) \xrightarrow{d} \frac{1/2\sigma_\varepsilon^2(W(1)^2 - 1)}{\int_0^1 W(r)^2 dr}$$

- Il s'agit d'une distribution non-standard complexe à manipuler
- En présence de non-stationnarité, ce type de distribution survient souvent

Théorie limite non standard des OLS : démonstration

- L'estimateur OLS est donnée par

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

- Si $\rho = 1$ et via une normalisation en n on a

$$n(\hat{\rho} - 1) = \frac{n^{-1} \sum_{t=1}^n X_{t-1} \varepsilon_t}{n^{-2} \sum_{t=1}^n X_t^2}$$

- A partir des lemmes (2) et (3) on obtient alors

$$n(\hat{\rho} - 1) \xrightarrow{d} \frac{1/2(W(1)^2 - 1)}{\int_0^1 W(r)^2 dr}$$

- On peut également démontrer (un peu plus fastidieux) que

$$\hat{t}_{\rho} = \frac{(\hat{\rho} - 1)}{\widehat{\sigma}_{\rho}} \xrightarrow{d} \frac{1/2(W(1)^2 - 1)}{\left(\int_0^1 W(r)^2 dr\right)^{1/2}}$$

En multipliant $\hat{\rho}$ par n de chaque côté on obtient

$$n(\hat{\rho} - \rho) = \frac{1}{n^{-1}} \left(n^{-1} \sum_t^n X_{t-1}^2 \right)^{-1} \left(n^{-1} \sum_t^n X_{t-1} \varepsilon_t \right)$$

Ce qui avec $\rho = 1$ nous donne

$$n(\hat{\rho} - 1) = \left(n^{-2} \sum_t^n X_{t-1}^2 \right)^{-1} \left(n^{-1} \sum_t^n X_{t-1} \varepsilon_t \right)$$

Théorie limite non standard des OLS : lemmes

Lemma (1)

Soit $X_t = \rho X_{t-1} + \varepsilon_t$, $\varepsilon_t \sim \text{i.i.d. } (0, \sigma_\varepsilon^2)$ et $X_0 = 0$. Si $\rho = 1$, alors

$$n^{-1/2} \sum_{t=1}^n \varepsilon_t \xrightarrow{d} \sigma_\varepsilon W(1)$$

- Voir le CLT de l'équation (7) pour la démonstration

Théorie limite non standard des OLS : lemmes

Lemma (2)

Soit $X_t = \rho X_{t-1} + \varepsilon_t$, $\varepsilon_t \sim \text{i.i.d. } (0, \sigma_\varepsilon^2)$ et $X_0 = 0$. Si $\rho = 1$, alors

$$n^{-2} \sum_{t=1}^n X_{t-1}^2 \xrightarrow{d} \sigma_\varepsilon^2 \int_0^1 W(r)^2 dr$$

- Voir (8) pour la démonstration

Théorie limite non standard des OLS : lemmes

Lemma (3)

Soit $X_t = \rho X_{t-1} + \varepsilon_t$, $\varepsilon_t \sim \text{i.i.d. } (0, \sigma_\varepsilon^2)$ et $X_0 = 0$. Si $\rho = 1$, alors

$$n^{-1} \sum_{t=1}^n X_{t-1} \varepsilon_t \xrightarrow{d} \frac{1}{2} \sigma_\varepsilon^2 (W(1)^2 - 1)$$

- Voir (9) pour la démonstration

Théorie limite non standard : Lemme 2

- Si $(t-1)/n \leq r_{t-1} < t/n$ on sait que

$$W_n(r) = n^{-1/2} X_{t-1} \sigma_{\mathbb{T}(X)}^{-1} \Rightarrow \sum_{t=1}^n W_n(r) = n^{-1/2} \sigma_{\mathbb{T}(X)}^{-1} \sum_{t=1}^n X_{t-1}$$

- On en déduit alors que $n^{-2} \sum_t^n X_{t-1}^2 = n^{-1} \sigma_{\mathbb{T}(X)}^2 \sum_{t=1}^n W_n(r)^2$

- On sait que $W_n(r)$ est constant si $(t-1)/n \leq r_{t-1} < t/n$ et donc

$$n^{-1} \sum_{t=1}^n W_n(r)^2 = \sum_{t=1}^n \int_{(t-1)/n}^{t/n} W_n(r)^2 dr = \int_0^1 W_n(r)^2 dr$$

- Le théorème de Mann-Wald fonctionnel nous assure alors que

$$n^{-2} \sum_t^n X_{t-1}^2 \xrightarrow{d} \sigma_{\mathbb{T}(X)}^2 \int_0^1 W(r)^2 dr \quad (8)$$

avec $\sigma_{\mathbb{T}(X)} = \sigma_\varepsilon$ (voir équation 6)

A partir de

$$\sum_{t=1}^n W_n(r) = n^{-1/2} \sigma_{\mathbb{T}(X)}^{-1} \sum_{t=1}^n X_{t-1}$$

on constate que

$$\sum_{t=1}^n X_{t-1} = n^{1/2} \sigma_{\mathbb{T}(X)}^1 \sum_{t=1}^n W_n(r)$$

en passant X_{t-1} au carré et en multipliant de chaque côté par n^{-2} on a

$$n^{-2} \sum_t^n X_{t-1}^2 = n^{-1} \sigma_{\mathbb{T}(X)}^2 \sum_{t=1}^n W_n(r)^2$$

Théorie limite non standard : Lemme 3

- Si $\rho = 1$ dans X_t alors $X_t^2 = (X_{t-1} + \varepsilon_t)^2 = X_{t-1}^2 + 2X_{t-1}\varepsilon_t + \varepsilon_t^2$
- Il vient immédiatement que

$$X_{t-1}\varepsilon_t = \frac{1}{2}(X_t^2 - X_{t-1}^2 - \varepsilon_t^2) \Rightarrow \sum_{t=1}^n X_{t-1}\varepsilon_t = \frac{1}{2}\left(X_n^2 - X_0^2 - \sum_{t=1}^n \varepsilon_t^2\right)$$

- Sous l'hypothèse que $X_0 = 0$ et en multipliant tout par n^{-1} on a

$$n^{-1} \sum_{t=1}^n X_{t-1}\varepsilon_t = \frac{1}{2}\left(n^{-1}X_n^2 - n^{-1} \sum_{t=1}^n \varepsilon_t^2\right)$$

- Via le lemme (1), le premier terme nous donne

$$n^{-1}X_n^2 = (n^{-1/2} \sum_{t=1}^n \varepsilon_t)^2 \xrightarrow{d} \sigma_\varepsilon^2 W(1)^2$$

- Via le théorème de Kolmogorov (3), $n^{-1} \sum_{t=1}^n \varepsilon_t^2 \xrightarrow{a.s.} \sigma_\varepsilon^2$ et

$$n^{-1} \sum_{t=1}^n X_{t-1}\varepsilon_t \xrightarrow{d} \frac{1}{2}\sigma_\varepsilon^2(W(1)^2 - 1) \quad (9)$$

Notons que $W(1)^2$ a une distribution de χ^2 car $W(1) \sim \mathcal{N}(0, 1)$

Plan

- | | | | |
|----------|---------------------------|----------|-----------------------------|
| 1 | Faits stylisés et rappels | 5 | Théorie limite standard |
| 2 | Non-stationnarité | 6 | Théorie limite non-standard |
| 3 | Intégration fractionnaire | 7 | Les régressions factices |
| 4 | Rappels de Probabilités | 8 | Conclusion |

Le concept de régression factice

- La non-stationnarité n'a pas pour seule conséquence l'émergence de distributions limites non-standards
- Supposons que l'on cherche à estimer le modèle

$$Y_t = \beta X_t + \varepsilon_t$$

avec $\beta = 0$

- Si Y_t et X_t sont des marches aléatoires, l'estimateur **OLS n'est pas consistant**
 - $\Rightarrow \hat{\beta}$ converge vers une **variable aléatoire non-dégénérée**
- $\hat{\beta}$ ne pouvant révéler l'absence de relation entre Y_t et X_t on parle de **régression factice**

Théorie limite et régression factice

Theorem (12)

Soit Y_t et X_t des marches aléatoires indépendantes, $Y_t = Y_{t-1} + \eta_t$ et $X_t = X_{t-1} + \nu_t$ avec $\eta_t \perp \nu_t$. On considère la régression

$$Y_t = \beta X_t + \varepsilon_t$$

avec $\beta = 0$. Alors la théorie limite de l'estimateur OLS de β

$$\hat{\beta} = 0 + \left(n^{-1} \sum_t^n X_{t-1}^2 \right)^{-1} \left(n^{-1} \sum_t^n X_{t-1} Y_{t-1} \right)$$

nous donne après normalisation par n

$$\hat{\beta} - 0 \xrightarrow{d} \left(\sigma_\nu \int_0^1 W_X(r)^2 dr \right)^{-1} \left(\sigma_\eta \int_0^1 W_X(r) W_Y(r) dr \right)$$

et révèle donc l'inconsistance de $\hat{\beta}$ si $\beta = 0$

Théorie limite et régression factice : normalisation

- En multipliant $\hat{\beta}$ par n de chaque côté on obtient

$$n(\hat{\beta} - \beta) = \frac{1}{n^{-1}} \left(n^{-1} \sum_t^n X_{t-1}^2 \right)^{-1} \left(n^{-1} \sum_t^n X_{t-1} Y_{t-1} \right)$$

- Ce qui nous donne

$$n(\hat{\beta} - 0) = \left(n^{-2} \sum_t^n X_{t-1}^2 \right)^{-1} \left(n^{-1} \sum_t^n X_{t-1} Y_{t-1} \right)$$

\Rightarrow En faisant passer le facteur n du terme de gauche, à droite

$$\hat{\beta} = \left(n^{-2} \sum_t^n X_{t-1}^2 \right)^{-1} \left(n^{-2} \sum_t^n X_{t-1} Y_{t-1} \right)$$

- La normalisation par n avait fonctionné pour la marche aléatoire
- Ici, n va disparaître et annihiler la vitesse de convergence

Théorie limite et régression factice : démonstration

- En posant $r = r_{t-1} = (t-1)/n$, on sait que

$$n^{-1/2}X_{t-1}\sigma_\nu^{-1} = W_{X_n}(r) \text{ et } n^{-1/2}Y_{t-1}\sigma_\eta^{-1} = W_{Y_n}(r)$$

- Par le Lemme (2) on obtient

$$n^{-2} \sum_t^n X_{t-1}^2 \xrightarrow{d} \sigma_\nu^2 \int_0^1 W(r)^2 dr$$

- Puisque $\eta_t \perp \nu_t$, une application multivariée du TCLF donne

$$\begin{pmatrix} \sigma_\eta^2 & 0 \\ 0 & \sigma_\nu^2 \end{pmatrix}^{-1/2} n^{-1/2} \sum_{t=1}^{[nr]} \begin{pmatrix} \eta_t \\ \nu_t \end{pmatrix} \xrightarrow{d} \begin{pmatrix} W_Y(r) \\ W_X(r) \end{pmatrix} \quad (10)$$

- Nous pouvons alors à analyser le dénominateur $n^{-2} \sum_t^n X_{t-1} Y_{t-1}$

Théorie limite et régression factice : démonstration

■ Sachant

- le Lemme (1)
- que $\eta_t \perp \nu_t$ et l'équation (10)
- que $W_n(r)$ est constant si $(t-1)/n \leq r_{t-1} < t/n$
- le théorème de Mann-Wald fonctionnel

$$\begin{aligned}
 n^{-2} \sum_t^n X_{t-1} Y_{t-1} &= n^{-1} \sum_{t=1}^n \sigma_\eta W_{Y_n}(r) \sigma_\nu W_{X_n}(r) \\
 &= \sigma_\eta \sigma_\nu \sum_{t=1}^n \int_{(t-1)/n}^{t/n} W_{Y_n}(r) W_{X_n}(r) dr \\
 &= \sigma_\eta \sigma_\nu \int_0^1 W_{Y_n}(r) W_{X_n}(r) dr \\
 &\xrightarrow{d} \sigma_\eta \sigma_\nu \int_0^1 W_Y(r) W_X(r) dr
 \end{aligned}$$

- n^{-1} disparaît pour faire apparaître l'intégrale

On a besoin d'un n^{-1} pour le TCLF multivarié et du dernier n^{-1} pour faire apparaître l'intégrale

Théorie limite et régression factice : démonstration

- On constate alors que

$$\begin{aligned}\hat{\beta} &= \left(n^{-2} \sum_t^n X_{t-1}^2 \right)^{-1} \left(n^{-2} \sum_t^n X_{t-1} Y_{t-1} \right) \\ &= \left(\sigma_\nu^2 \int_0^1 W_X(r)^2 dr \right)^{-1} \left(\sigma_\eta \sigma_\nu \int_0^1 W_Y(r) W_X(r) dr \right) \\ &\xrightarrow{d} \left(\sigma_\nu \int_0^1 W_X(r)^2 dr \right)^{-1} \left(\sigma_\eta \int_0^1 W_X(r) W_Y(r) dr \right)\end{aligned}$$

- Phillips (1986) démontre un résultat similaire pour $n^{-1/2} \hat{t}_\beta$

\Rightarrow Si $n \rightarrow \infty$ la probabilité de trouver $\hat{\beta}$ significatif approche 1 car la distribution de \hat{t}_β diverge à une vitesse $n^{1/2}$

- Phillips (1986) montre également qu'en présence d'une constante α dans la régression, la distribution de $\hat{\alpha}$ diverge

Pour être plus précis, Phillips (1986) trouve

$$n^{-1/2}\hat{\alpha} = \sigma_{\eta} \left(\int_0^1 W_Y(r)dr - \hat{\beta} \frac{\sigma_{\nu}}{\sigma_{\eta}} \int_0^1 W_X(r)dr \right)$$

On voit effectivement que la distribution de $\hat{\alpha}$ est non-dégénérée et diverge à la vitesse $n^{1/2}$ quand $n \rightarrow \infty$ car

$$\hat{\alpha} = n^{1/2} \sigma_{\eta} \left(\int_0^1 W_Y(r)dr - \hat{\beta} \frac{\sigma_{\nu}}{\sigma_{\eta}} \int_0^1 W_X(r)dr \right)$$

Régression factice et intégration fractionnaire

- Tsay et Chung (2000) étendent les résultats de Phillips (1986) au cas fractionnaire

- Soit deux processus à mémoire longue

$$y_t \sim I(\delta_y) \text{ et } x_t \sim I(\delta_x) \text{ avec } \delta_y, \delta_x \in (0, 1/2)$$

- Tsay et Chung (2000) montrent que

- le risque de régression factice existe dès lors que $\delta_y + \delta_x > 1/2$

\Rightarrow même si les deux processus sont stationnaires !

- La démonstration de ce résultat est bien plus complexe et dépasse de loin le niveau M2

Plan

- | | | | |
|----------|---------------------------|----------|-----------------------------|
| 1 | Faits stylisés et rappels | 5 | Théorie limite standard |
| 2 | Non-stationnarité | 6 | Théorie limite non-standard |
| 3 | Intégration fractionnaire | 7 | Les régressions factices |
| 4 | Rappels de Probabilités | 8 | Conclusion |

Ce qu'il faut retenir

- La notion de non-stationnarité est protéiforme
- Dans un cadre linéaire, la stationnarité disparaît si $\gamma(h)^2 = \infty$
- Il existe des processus stationnaires dont $\gamma(h)^2 < \infty$ mais $|\gamma(h)| = \infty$, on parle de processus mémoire longue
- Plus généralement les processus sont $I(\delta \in \mathbb{R})$ et $\delta = 0$ ou $\delta = 1$ sont des cas particuliers
- Si $\delta \geq 1/2$, la non-stationnarité survient car $\gamma(h)^2 = \infty$
- Si $\delta = 0$ on connaît la théorie limite des OLS
- Si $\delta = 1$ la théorie limite des OLS devient non standard et le risque de régression factice émerge
- Ce risque émerge en réalité pour $\delta_x + \delta_y > 1/2$

- Dahlhaus, R., Rao, S. S. (2006). Statistical inference for time-varying ARCH processes. *The Annals of Statistics*, 34(3), 1075-1114.
- Granger, C. W., Joyeux, R. (1980). An introduction to long-memory time series models and fractional differencing. *Journal of time series analysis*, 1(1), 15-29.
- Hosking, J. R. (1981). Asymptotic distributions of the sample mean, autocovariances, and autocorrelations of long-memory time series. *Journal of Econometrics*, 73(1), 261-284.
- Lieberman, O., Phillips, P.C.B. (2008). A complete asymptotic series for the autocovariance function of a long memory process. *Journal of Econometrics* 147, 99-103.
- Phillips, P.C.B. (1986). Understanding spurious regressions in econometrics. *Journal of Econometrics* 33, 311-340.
- Rao, S. S. (2006). On some nonstationary, nonlinear random processes and their stationary approximations. *Advances in Applied Probability*, 38(4), 1155-1172.
- Tsay, W.-J., Chung, C.-F. (2000). The spurious regression of fractionally integrated processes. *Journal of Econometrics*, 96(1), 155-182.