

# Econométrie des Séries Temporelles Univariées

## Chapitre 3 : les processus non stationnaires

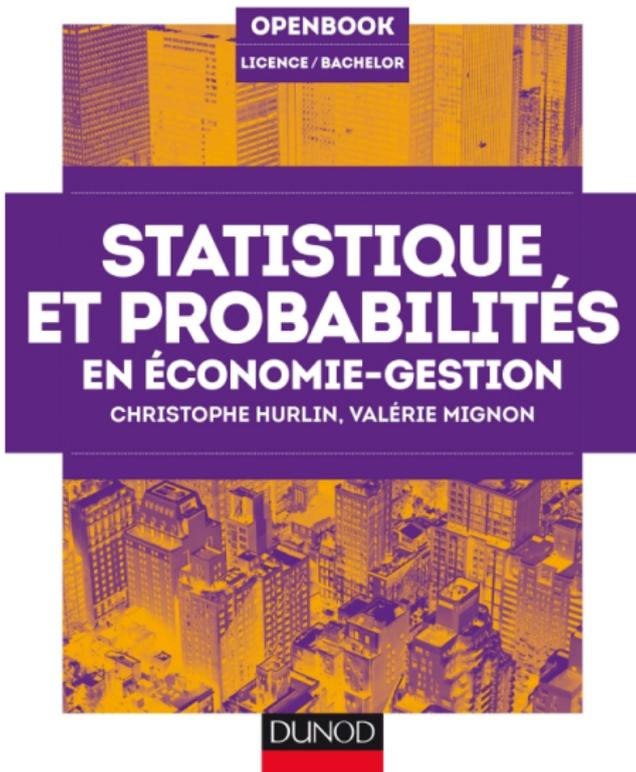
Gilles de Truchis

Master 1 ESA

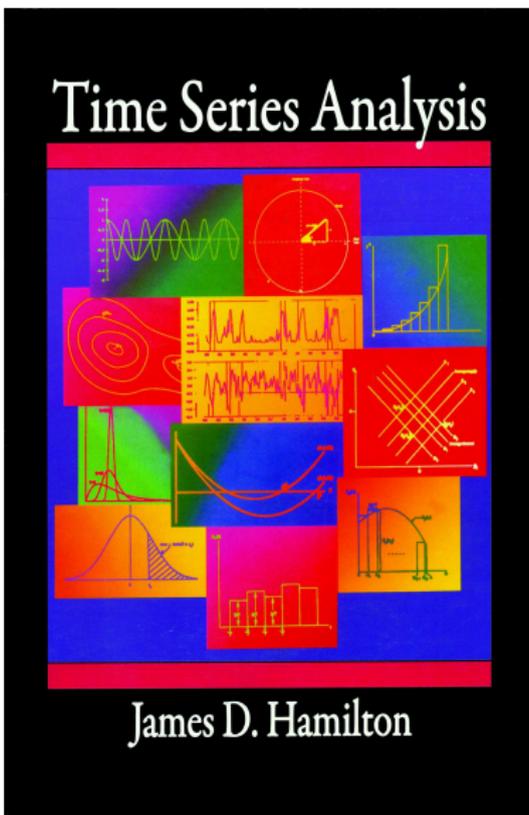
# Plan du chapitre

- 1 Faits stylisés et rappels
- 2 DS versus TS
- 3 Intégration fractionnaire
- 4 ARMA Saisonniers
- 5 Théorie limite standard
- 6 Théorie limite non-standard
- 7 Tests de racine unitaire
- 8 Les régressions factices
- 9 Conclusion
- 10 Rappels de Probabilités

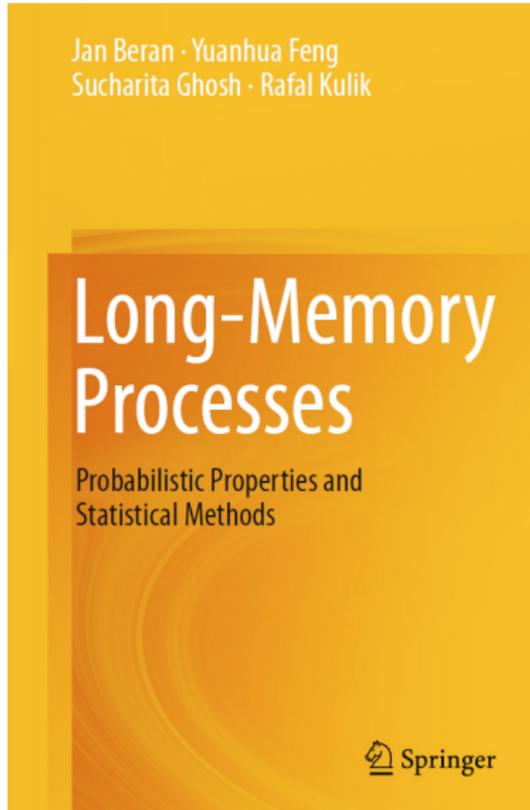
# Références



# Références



# Références



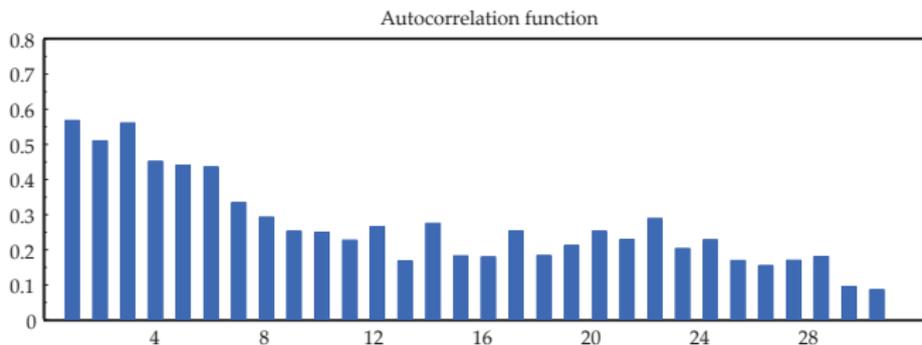
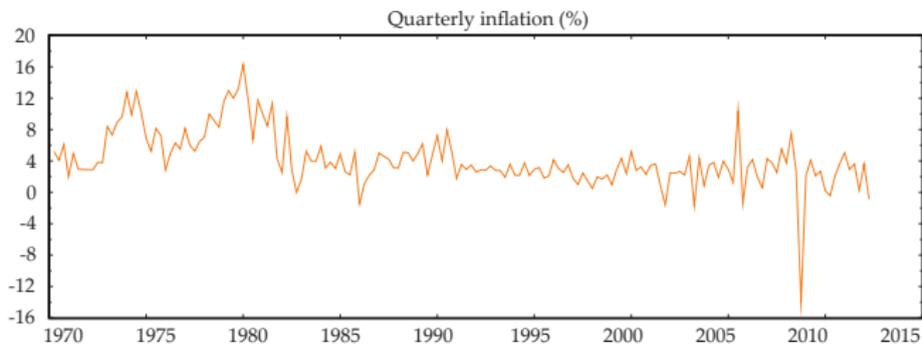
# Plan

- 1 Faits stylisés et rappels
- 2 DS versus TS
- 3 Intégration fractionnaire
- 4 ARMA Saisonniers
- 5 Théorie limite standard
- 6 Théorie limite non-standard
- 7 Tests de racine unitaire
- 8 Les régressions factices
- 9 Conclusion
- 10 Rappels de Probabilités

# Non-linéarité et cycles

- 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 ?

# Variables macroéconomiques



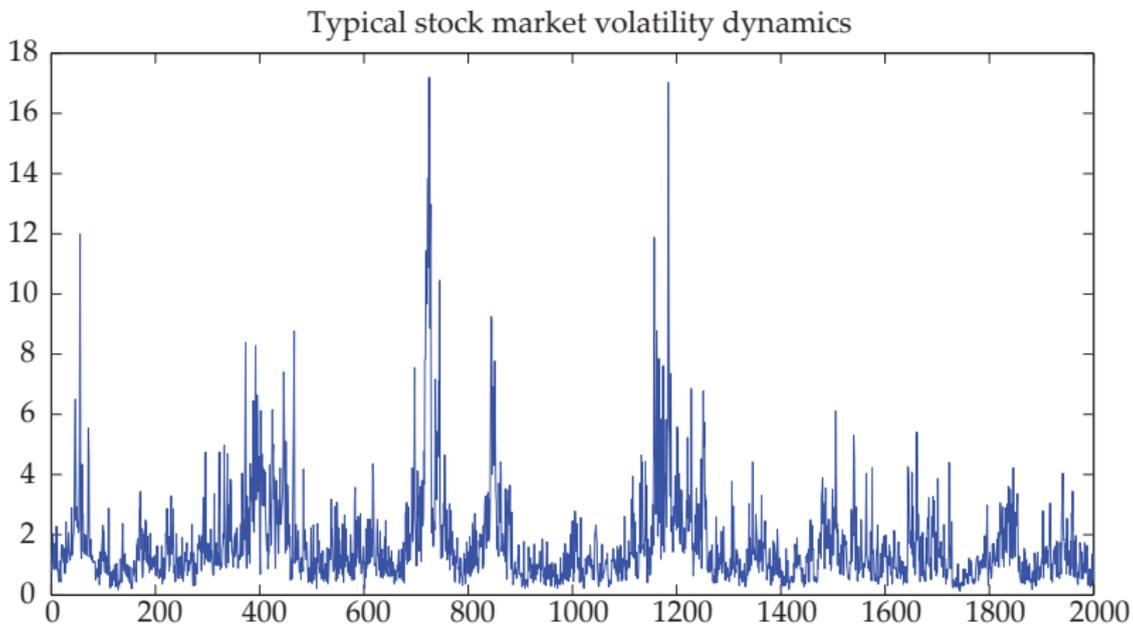
Quarterly US consumer price inflation 1970:1–2012:2 and its autocorrelation function

# Dynamique persistante

- 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 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

# Variables financières



# Dynamique extrême

- Difficile de déterminer visuellement si la volatilité est non-stationnaire
  - La série est animée par de nombreux événements extrêmes
  - 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

# Définir la non-stationnarité

- Plusieurs **phénomènes** peuvent mener à la **non-stationnarité**
  - ⇒ tendance déterministe
  - ⇒ saisonnalité
  - ⇒ tendance stochastique
  - ⇒ non-linéarités (dépassé le niveau M1)
  - ⇒ chocs de grandes valeurs et non-causalité (dépassé le niveau M1)

# Plan

- 1 Faits stylisés et rappels
- 2 DS versus TS
- 3 Intégration fractionnaire
- 4 ARMA Saisonniers
- 5 Théorie limite standard
- 6 Théorie limite non-standard
- 7 Tests de racine unitaire
- 8 Les régressions factices
- 9 Conclusion
- 10 Rappels de Probabilités

## 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  $\mathbb{E}(X_t) = \mu + \delta t$
- ⇒ 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 déterministe et estimation

- Soit  $X_t = \phi_0 + \delta t + \phi_1 X_{t-1} + \varepsilon_t$  un AR(1) avec tendance et  $|\phi_1| < 1$
- Malgré sa non-stationnarité, ce modèle peut s'estimer par OLS car

$$Z_t = \left( \phi_0 - \delta \frac{\phi_1}{1 - \phi_1} \right) + \phi_1 Z_{t-1} + \varepsilon_t$$

est une réécriture, valide si  $|\phi_1| < 1$ , qui s'obtient en considérant

$$Y_t = X_t - \delta \left( \sum_{i=1}^{t-1} \phi_1^i \right) \times t$$

i.e. l'écart à la tendance, dont on prendra l'espérance de la limite

⇒ la théorie limite reste standard et les tests usuels s'appliquent

Note Cela peut être généralisé au cas d'un AR( $p$ )

# Tendance déterministe et prévision

- La tendance affecte trivialement l'espérance conditionnelle

$$\hat{X}_{n+h} = \mathbb{E}(X_{n+h} | X_n, \dots, X_1) = \delta(n)_{n+h} + \mathbb{E}(\tilde{X}_{n+h} | X_n, \dots, X_1)$$

où  $\delta(n)_{n+h}$  est la prévision de la tendance pour  $n+h$  et

$$\tilde{X}_t = X_t - \delta(t)$$

- Par sa nature déterministe, la tendance n'affecte pas la variance

$$\mathbb{V}(X_{n+h} | X_n, \dots, X_1) = \mathbb{V}(\tilde{X}_{n+h} | X_n, \dots, X_1)$$

## 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 a  $X_t = X_0 + \sum_{j=0}^{t-1} \varepsilon_{t-j}$  (supposons  $X_0 = 0$ ) 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

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

- 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**  $\mu$

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

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

# Non-stationnarité globale

- 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

# Non-stationnarité locale

- Les processus dont les paramètres évoluent dans le temps sont possiblement **localement non-stationnaires**
  - Par exemple, un modèle à changement de régime peut être
    - stationnaire dans un régime
    - non-stationnaire dans autre régime (localement non-stationnaire)
- ⇒ globalement stationnaire ou non-stationnaire
- Ce type de processus étant non-linéaire ⇒ cours d'économétrie non-linéaire

# 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** (dépassé le niveau M2)
- Références : Rao (2006), Dahlhaus et Rao (2006)

# Plan

- 1 Faits stylisés et rappels
- 2 DS versus TS
- 3 Intégration fractionnaire
- 4 ARMA Saisonniers
- 5 Théorie limite standard
- 6 Théorie limite non-standard
- 7 Tests de racine unitaire
- 8 Les régressions factices
- 9 Conclusion
- 10 Rappels de Probabilités

## ARMA stationnaire

- Au Chapitre 1 nous avons introduit les ARMA leurs propriétés
- Supposons  $X_t \sim ARMA(p, q)$

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

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

- On sait que sous cette condition de stationnarité

$$\sum_{i=0}^{\infty} |\psi_j| < \infty$$

## Rappel sur l'ergodicité des ARMA

- Or, à partir de la forme  $MA(\infty)$  de  $X_t$  on a

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

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

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

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

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

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

et la convolution de deux séries absolument sommables est absolument sommable

# Integration

## Definition (1)

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

## Definition (2)

Si un processus est **faiblement dépendant** après  $\delta$  différenciations, il est dit **intégré d'ordre  $\delta$**  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  $\Delta X_t \sim I(0)$

## ARIMA

## Definition (3)

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

$$\phi(L)\Delta^\delta X_t = \alpha + \theta(L)\varepsilon_t, \quad \varepsilon_t \sim i.i.d.(0, \sigma_\varepsilon^2 < \infty)$$

- Exemple : en cas de racine unitaire on a  $\delta = 1$  et donc

$$\phi(L)\Delta X_t = \alpha + \theta(L)\varepsilon_t \Leftrightarrow \Delta X_t = \mu + \psi(L)\varepsilon_t$$

avec  $\mu = \phi(L)^{-1}\alpha$  et  $\psi(L) = \phi(L)^{-1}\theta(L)$

- Dans le même esprit que la marche aléatoire on observe que

$$X_t = X_0 + \mu t + \psi(L) \sum_{j=1}^{\infty} \varepsilon_{t-j}$$

$\Rightarrow$  Dans les cas simples  $\delta \in \mathbb{Z}$ , mais que se passe-t-il si  $\delta \in \mathbb{R}$  ?

# 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  $(\delta)_j \times \Gamma(j + 1)^{-1} L^j$

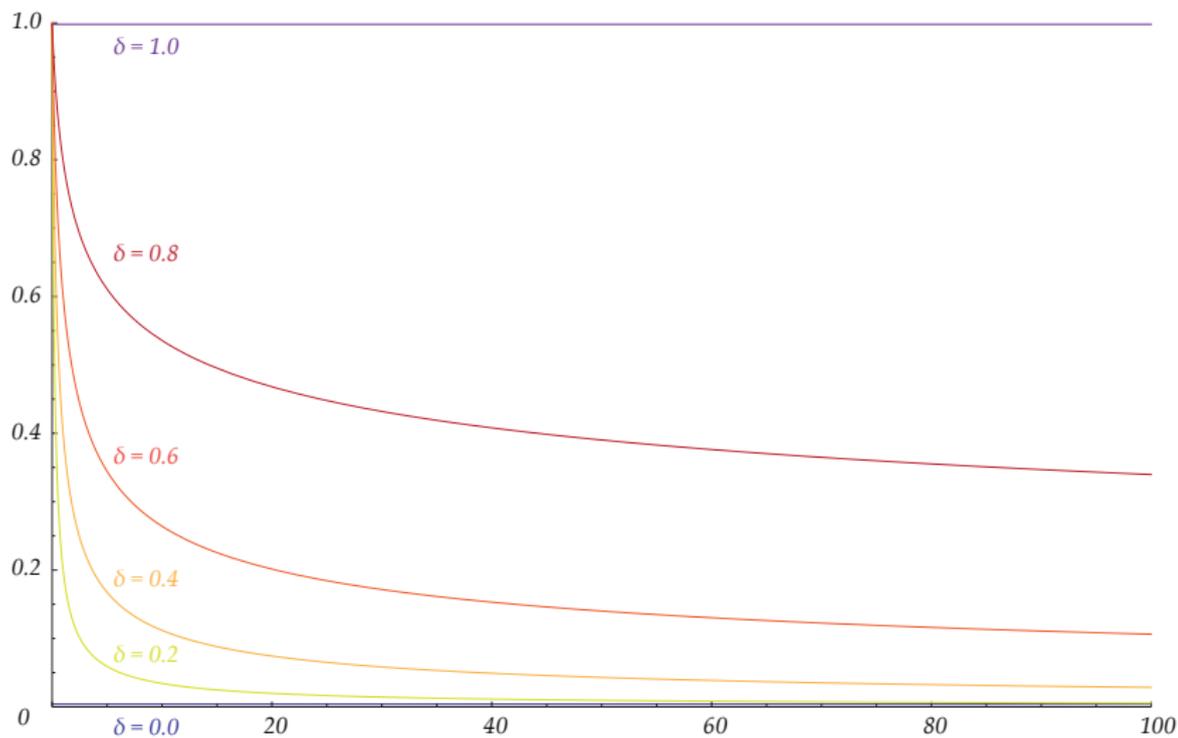
Rappelons que la fonction Gamma est telle que

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

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

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

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

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

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

⇒  $j^{2(\delta-1)=p<-1}$  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 (4)

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  $h \rightarrow \infty$

# ARFIMA

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

## Definition (5)

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

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

Note 1 Les démonstrations précédentes tiennent pour un ARFIMA bien que les formules se complexifient

Note 2 Pour  $\delta = 1$ , une racine unitaire est présente et on parle d'ARIMA

# ARFIMA

- A la différence des processus ARIMA( $p, \delta, q$ )

- $\Rightarrow X_t$  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-persistent si  $\delta < 0$

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

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

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

# Plan

- 1 Faits stylisés et rappels
- 2 DS versus TS
- 3 Intégration fractionnaire
- 4 ARMA Saisonniers
- 5 Théorie limite standard
- 6 Théorie limite non-standard
- 7 Tests de racine unitaire
- 8 Les régressions factices
- 9 Conclusion
- 10 Rappels de Probabilités

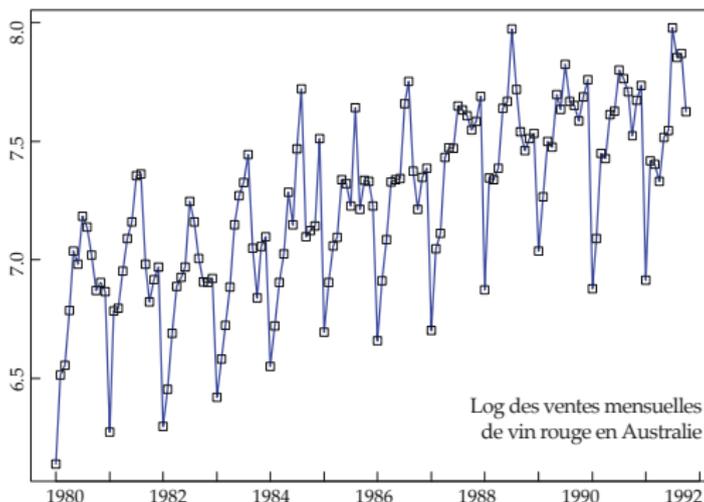
# Saisonnalité

■ La saisonnalité stochastique n'implique pas la non-stationnarité

⇒ mais s'analyse à travers un modèle général : Seasonal ARIMA

e.g. les ventes mensuelles de vin montrent une forte saisonnalité

⇒ pic en juillet et creux en janvier en plus d'une tendance haussière



## Dépendance saisonnière

- La dépendance au passé sera donc importante à 12 mois

$$\rho_{12}, \rho_{24}, \rho_{36}, \dots$$

entre juillet (janvier) de l'année en cours et l'année passée

⇒ Un modèle ARMA( $P, Q$ ) $_s$  troué de **span**  $s = 12$  sera approprié

e.g. 1 dans le cas d'un MA(1) $_{12}$  on aura

$$X_t = (1 + \vartheta_1 L^{12})\varepsilon_t \Leftrightarrow X_t = \varepsilon_t + \vartheta_1 \varepsilon_{t-12}$$

e.g. 2 dans le cas d'un AR(1) $_{12}$  on aura

$$(1 - \varphi_1 L^{12})X_t = \varepsilon_t \Leftrightarrow X_t = \varphi_1 X_{t-12} + \varepsilon_t$$

Note Le span dépend de la fréquence :

$$s = 2 \text{ (semestriel)}, s = 4 \text{ (trimestriel)}, s = 52 \text{ (hebdomadaire)}$$

## Filtre saisonnier et SARMA

- Après application d'un  $\text{ARMA}(P, Q)_s$ , une dépendance peut rester
- L' $\text{ARMA}(P, Q)_s$  agit comme un filtre qui retourne  $Y_t$

$$\Phi(L)_s X_t = \Theta(L)_s Y_t$$

- Un second filtre de type  $\text{ARMA}(p, q)$  est donc nécessaire

$$\Phi(L) Y_t = \Theta(L) \varepsilon_t$$

afin que la dépendance disparaisse et de récupérer  $\varepsilon_t \sim WN(0, \sigma_\varepsilon^2)$

$\Rightarrow$  Le produit des deux filtres donne le modèle  $\text{SARMA}(p, q) \times (P, Q)_s$

$$\Phi(L)\Phi(L)_s X_t = \Theta(L)\Theta(L)_s \varepsilon_t$$

Note Les conditions de stationnarité et inversibilité usuelles s'appliquent

SARMA(0, 0)  $\times$  (0, 1)<sub>4</sub>

- Il s'agit d'un SMA du type

$$X_t = \varepsilon_t + \vartheta_1 \varepsilon_{t-4}$$

approprié pour des données trimestrielles par exemple

Type MA( $q$ )  $\Rightarrow$  inversible si ses 4 racines

$$-i\sqrt[4]{\frac{1}{\vartheta_1}}, \quad i\sqrt[4]{\frac{1}{\vartheta_1}}, \quad \sqrt[4]{\frac{1}{\vartheta_1}}, \quad -\sqrt[4]{\frac{1}{\vartheta_1}}$$

sont en dehors du cercle unitaire et donc  $|\vartheta_1| < 1$

- Si  $|\vartheta_1| < 1$ ,  $X_t$  possèdent une forme AR( $\infty$ )

$$X_t = - \sum_{j=1}^{\infty} \vartheta_1^j X_{t-4j} + \varepsilon_t$$

- On peut aisément vérifier que  $\gamma_0 = (1 + \vartheta_1^2)\sigma_\varepsilon^2$ ,

$$\gamma_4 = \vartheta_1 \sigma_\varepsilon^2$$

et  $\gamma_j = 0$  pour  $j \neq 0, 4$

## SARMA(0, 1) $\times$ (0, 1)<sub>5</sub>

- Il s'agit d'un SARMA du type

$$X_t = (1 + \theta_1 L)(1 + \vartheta_1 L^5)\varepsilon_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \vartheta_1 \varepsilon_{t-5} + \theta_1 \vartheta_1 \varepsilon_{t-6}$$

approprié pour des données journalières (jours ouvrés)

Type MA(6)  $\Rightarrow$  inversible si ses 6 racines

... sont en dehors du cercle unitaire

- On peut aisément vérifier que

$$\gamma_0 = (1 + \theta_1^2 + \vartheta_1^2 + \theta_1^2 \vartheta_1^2)\sigma_\varepsilon^2 = (1 + \theta_1^2)(1 + \vartheta_1^2)\sigma_\varepsilon^2$$

$$\gamma_1 = (\theta_1 + \theta_1 \vartheta_1^2)\sigma_\varepsilon^2 = \theta_1(1 + \vartheta_1^2)\sigma_\varepsilon^2$$

$$\gamma_4 = (\theta_1 \vartheta_1)\sigma_\varepsilon^2$$

$$\gamma_5 = (\vartheta_1 + \theta_1^2 \vartheta_1)\sigma_\varepsilon^2 = \vartheta_1(1 + \theta_1^2)\sigma_\varepsilon^2$$

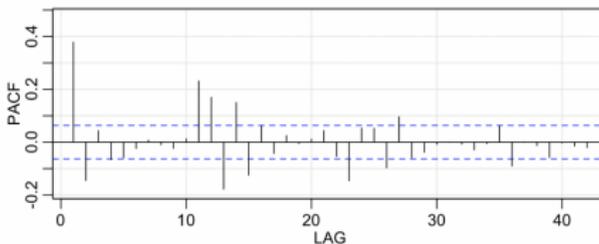
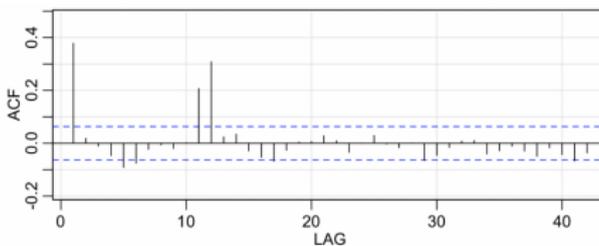
$$\gamma_6 = (\theta_1 \vartheta_1)\sigma_\varepsilon^2$$

et  $\gamma_j = 0$  pour  $j \neq 0, 1, 4, 5, 6$  et  $j > 6$

# SARMA(0, 1) $\times$ (0, 1)<sub>12</sub>

- Etudions le comportement (simulé) des ACFs et PACFs si

$$\theta_1 = 0.7 \text{ et } \vartheta_1 = 0.6 \text{ et } \sigma_\varepsilon^2 = 1$$



# SARMA(0, $q$ ) $\times$ (0, $Q$ ) $_s$

- Quelques généralités se dégagent de l'exemple précédent
  - L'ACovF et l'ACF sont atypiques autour de  $s$
- $\Rightarrow$  Autour de  $s$ , les  $q$  autocovariances sont symétriques
- Comme pour tout MA( $q$ ), la PACF ne s'annule pas
- $\Rightarrow$  Elle décroît vers 0 quand  $h \rightarrow \infty$
- $\Rightarrow$  Des résurgences apparaîtront autour  $s, 2s, 3s, \dots$

# SARMA(0, 1) $\times$ (1, 0)<sub>7</sub>

- Il s'agit d'un SARMA du type

$$(1 - \varphi_1 L^7)X_t = (1 + \theta_1 L)\varepsilon_t$$

approprié pour des données journalières (semaine pleine)

Type AR(7) + MA(1)  $\Rightarrow$  stationnaire et inversible si les 7 racines

... de  $(1 - \varphi_1 L^7)$  sont en dehors du cercle unitaire et si  $|\theta_1| < 1$

- On peut vérifier que

$$\gamma_0 = (1 + \theta_1^2)(1 - \varphi_1^2)^{-1}\sigma_\varepsilon^2$$

$$\gamma_1 = \varphi_1 \mathbb{E}(X_{t-7}X_{t-1}) + \mathbb{E}(\varepsilon_t X_{t-1}) - \theta_1 \mathbb{E}(\varepsilon_{t-1} X_{t-1}) = \varphi_1 \gamma_6 - \theta_1 \sigma_\varepsilon^2$$

$$\gamma_j = \varphi_1 \gamma_{j-7}, \quad \forall j > 2$$

$$\gamma_6 = \varphi_1 \gamma_1 \Rightarrow \gamma_1 = -\theta_1 \sigma_\varepsilon^2 (1 - \varphi_1^2)^{-1}$$

$$\gamma_k = \varphi_1^j \gamma_1, \quad \forall k = 7j \pm 1, \quad j = 1, 2, \dots \text{ et}$$

$$\gamma_k = \varphi_1^j \gamma_0, \quad \forall k = 7j, \quad j = 0, 1, 2, \dots$$

et  $\gamma_k = 0$  pour les  $k$  non considérés ci-dessus

SARMA(1, 1)  $\times$  (2, 1)<sub>12</sub>

- Ici les filtres sont multiplicatifs mais il existe des filtres additifs

⇒ Cela va impacter l'écriture des polynômes

e.g. L'écriture d'un SARMA(1, 1)  $\times$  (2, 1)<sub>12</sub> sera

$$(1 - \varphi_1 L)(1 - \phi_1 L^{12} - \phi_2 L^{24})X_t = (1 + \theta_1 L)(1 + \vartheta_1 L^{12})\varepsilon_t$$

e.g. L'écriture d'un SARMA(1, 1)  $+$  (2, 1)<sub>12</sub> sera

$$(1 - \varphi_1 L - \phi_1 L^{12} - \phi_2 L^{24})X_t = (1 + \theta_1 L + \vartheta_1 L^{12})\varepsilon_t$$

## SARIMA( $p, \delta, q$ ) $\times$ ( $P, D, Q$ ) $_s$

- On a vu que  $\Delta^\delta$  était important pour stationnariser une série
- On peut également construire un opérateur

$$\Delta_s^D = (1 - L^s)^D$$

de différenciation saisonnière

- En effet, certaines séries sont cyclo-stationnaires et avec  $D = 1$ , si

$$\Delta_s X_t = X_t - X_{t-s}$$

est stationnaire, c'est que  $X_t$  était une marche aléatoire saisonnière

e.g.  $X_{01/2021} = X_{01/2020} + \varepsilon_{01/2021}$  suit une marche aléatoire saisonnière

- Plus généralement,  $X_t \sim \text{SARIMA}(p, \delta, q) \times (P, D, Q)_s$  si

$$Y_t = \Delta^\delta \Delta_s^D X_t \sim \text{SARMA}(p, q) \times (P, Q)_s$$

# X13-ARIMA-SEATS

- La filtration de la saisonnalité est une tâche fastidieuse
- Des procédures de traitement automatique existent depuis 1970

⇒ X11-ARIMA, TRAMO-SEATS, X12-ARIMA (dispo sous SAS)

Réf. Une revue de ses méthodes est faite par Darné (2004)

- La plus récente et populaire : la procédure X13-ARIMA-SEATS

⇒ basée sur des régressions de SARIMA, elle peut aussi traiter

... des effets très spécifiques comme

- valeurs aberrantes
- jours ouvrables
- effets de Pâques
- traitement spectral de la saisonnalité résiduelle
- ...

# Plan

- 1 Faits stylisés et rappels
- 2 DS versus TS
- 3 Intégration fractionnaire
- 4 ARMA Saisonniers
- 5 Théorie limite standard
- 6 Théorie limite non-standard
- 7 Tests de racine unitaire
- 8 Les régressions factices
- 9 Conclusion
- 10 Rappels de Probabilités

# 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 (7),  $\sum \sigma_X^2 \sigma_\varepsilon^2 / t^2 < \infty$  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** (10) 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 :  $\widehat{\beta} \xrightarrow{p} \beta$
- Avec un transformation  $\mathbb{T}(\widehat{\beta}) \xrightarrow{d} \mathcal{L}$  on peut montrer que

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

- 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** (9) 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** (11) on a alors

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

## 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** (11) on a alors

$$\sqrt{n}(\hat{\rho} - \rho) \xrightarrow{d} \frac{\mathcal{N}(0, \sigma_\varepsilon^2 \sigma_X^2)}{\sigma_X^2} \stackrel{d}{=} \mathcal{N}(0, 1 - \rho^2) \quad (3)$$

- 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  $\hat{\rho}$

# Plan

- 1 Faits stylisés et rappels
- 2 DS versus TS
- 3 Intégration fractionnaire
- 4 ARMA Saisonniers
- 5 Théorie limite standard
- 6 Théorie limite non-standard
- 7 Tests de racine unitaire
- 8 Les régressions factices
- 9 Conclusion
- 10 Rappels de Probabilités

## 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 (3) puisque

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

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

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

⇒ 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

$$\sigma_{\mathbb{T}(X)}^2 = \mathbb{V}(\mathbb{T}(X_n)) = \mathbb{E} \left( n^{-1} \left( \sum_{t=1}^n \varepsilon_t \right)^2 \right) = n^{-1} n \sigma_\varepsilon^2 = \sigma_\varepsilon^2 \quad (4)$$

- D'après le TCL de Lindeberg-Lévy (8), 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

$$W_n(1) = n^{-1/2} \sum_{s=1}^n \varepsilon_s \sigma_{\mathbb{T}(X)}^{-1} \tag{5}$$

# Les processus de Wiener

- $W_n(1) \xrightarrow{d} W(1)$  avec  $W(1)$  un processus de Wiener

## Definition (6)

Soit un espace probabilisé  $(\Omega, \mathcal{F}, \mathbb{P})$ .  $W : \Omega \times [0, 1] \rightarrow \mathbb{R}^1$  est un processus de Wiener standard si pour chaque  $r \in [0, 1]$ ,  $W(r)$  est  $\mathcal{F}$ -mesurable et si

- (a) le processus  $W(r)$  débute en 0 :  $\mathbb{P}(W(0) = 0) = 1$
- (b) le processus  $W(r)$  possède des incréments indépendants : si  $0 \leq t_0 \leq t_1 \leq \dots \leq t_k \leq 1$ ,  $W(t_i) - W(t_{i-1})$  est indépendant de  $W(t_j) - W(t_{j-1})$  pour  $j = 1, \dots, k$ ,  $j \neq i, \forall i = 1, \dots, k$
- (a) le processus  $W(r)$  possède des incréments normalement distribués : pour  $0 \leq a \leq b \leq 1$ , l'incrément  $W(b) - W(a) \sim \mathcal{N}(0, b - a)$

- Mais vers quoi converge  $W_n(r)$  ?

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  $\varepsilon_t$ ) de  $r$

## Theorem (1 : Théorème de Donsker)

Soit  $\varepsilon_t$  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 (2 : Functional Continuous Mapping Theorem)

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

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

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

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

## Théorie limite non standard des OLS

## Theorem (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  $\rho = 1$ , la distribution limite de l'estimateur OLS est donnée par

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

- 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

$$\hat{t}_\rho = \frac{(\hat{\rho} - 1)}{\hat{\sigma}_\rho} \xrightarrow{d} \frac{1/2(W(1)^2 - 1)}{\left(\int_0^1 W(r)^2 dr\right)^{1/2}} \quad (6)$$

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  $X_t = \rho X_{t-1} + \varepsilon_t$ ,  $\varepsilon_t \sim$  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 (5) pour la démonstration

## Théorie limite non standard des OLS : lemmes

## Lemma (2)

Soit  $X_t = \rho X_{t-1} + \varepsilon_t$ ,  $\varepsilon_t \sim$  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 (7) pour la démonstration

## Théorie limite non standard des OLS : lemmes

## Lemma (3)

Soit  $X_t = \rho X_{t-1} + \varepsilon_t$ ,  $\varepsilon_t \sim$  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 (8) 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  $n^{-2} \sum_t X_{t-1}^2 = n^{-1} \sigma_{\mathbb{T}(X)}^2 \sum_{t=1}^n W_n(r)^2$
- 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

$$n^{-2} \sum_t X_{t-1}^2 \xrightarrow{d} \sigma_{\mathbb{T}(X)}^2 \int_0^1 W(r)^2 dr \quad (7)$$

avec  $\sigma_{\mathbb{T}(X)} = \sigma_\varepsilon$  (voir équation 4)

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

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

## 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 (6),  $n^{-1} \sum_{t=1}^n \varepsilon_t^2 \xrightarrow{a.s.} \sigma_\varepsilon^2$  et

$$n^{-1} \sum_{t=1}^n X_{t-1}\varepsilon_t \xrightarrow{d} \frac{1}{2} \sigma_\varepsilon^2 (W(1)^2 - 1) \quad (8)$$

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

# Plan

- 1 Faits stylisés et rappels
- 2 DS versus TS
- 3 Intégration fractionnaire
- 4 ARMA Saisonniers
- 5 Théorie limite standard
- 6 Théorie limite non-standard
- 7 Tests de racine unitaire
- 8 Les régressions factices
- 9 Conclusion
- 10 Rappels de Probabilités

# Généralité sur les tests de racine unitaire

- Soit un AR(1) dont on souhaite tester si  $\rho = 1$
- Les difficultés résident dans la formulation d'un test car
  - la distribution limite ne sera pas standard
- Dickey-Fuller ont proposé deux tests :  $DF$  et  $DF$  augmenté ( $ADF$ )
- Phillips-Perron ont proposé une alternative non-paramétrique d'ADF

## Un premier test de racine unitaire

- Soit  $x_t$  le taux de change journalier EUR/USD sur 2000-2014
- On sait que si  $x_t = \rho x_{t-1} + \varepsilon_t$  avec  $\rho = 1$ , d'après Th.3

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

- $\mathcal{L}(\hat{\rho})$  est tabulée au S71 et peut servir pour tester  $H_0 : \rho = 1$
- Calculons la statistique  $z_{\hat{\rho}}$  pour  $n = 3760$  sachant que

$$\hat{\rho} = 0.9999919311 \Rightarrow z_{\hat{\rho}} = n(\hat{\rho} - 1) = -0.03033893$$

- Dans la table, pour  $n > 500 \equiv \infty$ , 95% des fois lorsque le processus est vraiment une marche aléatoire,  $s_{\hat{\rho}}$  est supérieur à -8.1
- $\Rightarrow$  puisque  $-0.03 > -8.1$ , l'hypothèse nulle  $\rho = 1$  est acceptée au seuil de 5% et  $x_t \sim I(1)$

## Un premier test de racine unitaire : annexe

**TABLE B.5**  
**Critical Values for the Phillips-Perron  $Z_\alpha$  Test and for the Dickey-Fuller Test**  
**Based on Estimated OLS Autoregressive Coefficient**

Sample size <i>T</i>	Probability that $T(\hat{\rho} - 1)$ is less than entry							
	0.01	0.025	0.05	0.10	0.90	0.95	0.975	0.99
<i>Case 1</i>								
25	-11.9	-9.3	-7.3	-5.3	1.01	1.40	1.79	2.28
50	-12.9	-9.9	-7.7	-5.5	0.97	1.35	1.70	2.16
100	-13.3	-10.2	-7.9	-5.6	0.95	1.31	1.65	2.09
250	-13.6	-10.3	-8.0	-5.7	0.93	1.28	1.62	2.04
500	-13.7	-10.4	-8.0	-5.7	0.93	1.28	1.61	2.04
$\infty$	-13.8	-10.5	-8.1	-5.7	0.93	1.28	1.60	2.03
<i>Case 2</i>								
25	-17.2	-14.6	-12.5	-10.2	-0.76	0.01	0.65	1.40
50	-18.9	-15.7	-13.3	-10.7	-0.81	-0.07	0.53	1.22
100	-19.8	-16.3	-13.7	-11.0	-0.83	-0.10	0.47	1.14
250	-20.3	-16.6	-14.0	-11.2	-0.84	-0.12	0.43	1.09
500	-20.5	-16.8	-14.0	-11.2	-0.84	-0.13	0.42	1.06
$\infty$	-20.7	-16.9	-14.1	-11.3	-0.85	-0.13	0.41	1.04
<i>Case 4</i>								
25	-22.5	-19.9	-17.9	-15.6	-3.66	-2.51	-1.53	-0.43
50	-25.7	-22.4	-19.8	-16.8	-3.71	-2.60	-1.66	-0.65
100	-27.4	-23.6	-20.7	-17.5	-3.74	-2.62	-1.73	-0.75
250	-28.4	-24.4	-21.3	-18.0	-3.75	-2.64	-1.78	-0.82
500	-28.9	-24.8	-21.5	-18.1	-3.76	-2.65	-1.78	-0.84
$\infty$	-29.5	-25.1	-21.8	-18.3	-3.77	-2.66	-1.79	-0.87

The probability shown at the head of the column is the area in the left-hand tail.

Source: Wayne A. Fuller, *Introduction to Statistical Time Series*, Wiley, New York, 1976, p. 371.

## Le test de Dickey-Fuller : intuition

- Dans un premier temps, considérons un AR(1) simple

$$X_t = \rho X_{t-1} + \varepsilon_t$$

- Dickey-Fuller (DF) proposent la reformulation suivante

$$\Delta X_t = (\rho - 1)X_{t-1} + \varepsilon_t = \phi X_{t-1} + \varepsilon_t$$

- La formulation du test de racine unitaire devient alors

$$H_0 : \phi = 0 \text{ versus } H_1 : \phi < 0$$

- Grâce au Th.3, on connaît la distribution limite des OLS pour

$$\phi = (\rho - 1)$$

⇒ la distribution de la statistique de “type” Student

$$\hat{\tau}_3 = \frac{\hat{\phi}}{\hat{\sigma}_\phi}$$

ne sera pas standard non plus (voir Eq. 6)

## Le test de Dickey-Fuller : cas général

- Dans la pratique il convient d'envisager la présence de

$$\mu \neq 0 \text{ et/ou } \delta \neq 0$$

⇒ 3 modèles distinct sont alors envisageables

M1 Le modèle le plus général, avec  $\mu \neq 0$  et/ou  $\delta \neq 0$

$$\Delta X_t = \mu + \delta t + \phi_1 X_{t-1} + \varepsilon_t$$

M2 Le modèle sans tendance mais avec  $\mu \neq 0$

$$\Delta X_t = \mu + \phi_2 X_{t-1} + \varepsilon_t$$

M3 Le modèle le plus spécifique décrivant une simple marche aléatoire

$$\Delta X_t = \phi_3 X_{t-1} + \varepsilon_t$$

- Comme sous  $H_0$  on a une racine unitaire, les distributions de

$$\hat{\phi}_1, \hat{\phi}_2, \hat{\phi}_3$$

sont non-standards, tout comme les distributions de

$$\hat{\tau}_1 = \hat{\phi}_1 / \hat{\sigma}_\phi, \quad \hat{\tau}_2 = \hat{\phi}_2 / \hat{\sigma}_\phi, \quad \hat{\tau}_3 = \hat{\phi}_3 / \hat{\sigma}_\phi$$

## Le test de Dickey-Fuller : les alternatives

- L'existence de plusieurs alternatives complique l'implémentation

e.g. Supposons que l'observation des données révèle une tendance

⇒ la question est alors de savoir si cette tendance vient de

$$\Delta X_t = \mu + \phi X_{t-1} + \varepsilon_t \Leftrightarrow X_t = \mu + X_{t-1} + \varepsilon_t \text{ sous } H_0$$

i.e. une marche aléatoire avec dérive, ou si elle vient de

$$\Delta X_t = \mu + \delta t + \phi X_{t-1} + \varepsilon_t \Leftrightarrow X_t = \mu + \delta t + \rho X_{t-1} + \varepsilon_t \text{ sous } H_1$$

i.e. un AR(1) stationnaire ( $\phi < 0 \Leftrightarrow \rho < 1$ ) autour d'un trend

- Dans cette situation, il est préférable de tester  $\phi = 0$  via

$$\hat{\tau}_1 = \hat{\phi}_1 / \hat{\sigma}_\phi$$

et de compléter cela par un test de nullité jointe (type Fisher)

$$F_1 := \hat{\phi}_1 = \hat{\delta} = 0$$

ou

$$F_2 := \hat{\phi}_1 = \hat{\delta} = \hat{\mu} = 0$$

# Le test de Dickey-Fuller : méthodologie

- L'étape 1 de la procédure à suivre en absence d'à priori sera
  - 1.a Estimer M1 et tester  $H_0 : \phi_1 = 0$  avec  $\hat{\tau}_1$
  - 1.b Si rejet de  $H_0$ ,  $(X_t - \delta t) \sim I(0)$ 
    - $\Rightarrow$  on peut tester  $\delta = 0$  et  $\mu = 0$  avec des valeurs critiques usuelles
  - 1.c Si non-rejet de  $H_0$ ,  $X_t$  possède une racine unitaire
    - $\Rightarrow$  il faut tester  $\hat{\phi}_1 = \hat{\delta} = 0$  avec  $F_1$  (valeurs critiques tabulées)
  - 1.d Si rejet, on garde M1 et on recommence la procédure sur  $\Delta X_t$
  - 1.e Si non-rejet, on peut soit considérer  $F_2$ , soit passer à l'étape 2
    - $\Rightarrow$  Si non-rejet avec  $F_2$ , on retient M3

## Le test de Dickey-Fuller : méthodologie

- L'étape 2 de la procédure à suivre en absence d'à priori sera
  - 2.a Estimer M2 et tester  $H_0 : \phi_2 = 0$  avec  $\hat{\tau}_2$
  - 2.b Si rejet de  $H_0$ ,  $X_t \sim I(0)$ 
    - $\Rightarrow$  on peut tester  $\mu = 0$  avec des valeurs critiques usuelles
  - 2.c Si non-rejet de  $H_0$ ,  $X_t$  possède une racine unitaire
    - $\Rightarrow$  il faut tester  $\hat{\phi}_2 = \hat{\mu} = 0$  avec  $F_3$  (valeurs critiques tabulées)
  - 2.d Si rejet, on garde M2 et on recommence la procédure sur  $\Delta X_t$
  - 2.e Si non-rejet, on peut passer à l'étape 3
  - 3.a Estimer M3 et tester  $H_0 : \phi_3 = 0$  avec  $\hat{\tau}_3$
  - 3.b Si rejet de  $H_0$ ,  $X_t \sim I(0)$
  - 3.c Si non-rejet, on garde M3 on recommence la procédure sur  $\Delta X_t$

## Les valeurs critiques de Dickey-Fuller

- La forme non-standard des distributions limites de

$$\hat{\tau}_1, \hat{\tau}_2, \hat{\tau}_3, \hat{F}_1, \hat{F}_2, \hat{F}_3$$

empêchent le calcul analytique de valeurs critiques

- Dickey-Fuller proposent des valeurs critiques tabulées pour

$$F_i = \frac{(SSR_c - SSR_{nc})/r}{SSR_{nc}/(n - k)}$$

avec  $r$  et  $k$  les # de restrictions et paramètres (modèle non-contraint)

	Étape (1)						Étape (2)		
	$\hat{F}_2 := \hat{\phi}_1 = \hat{\mu} = \hat{\delta} = 0$			$\hat{F}_1 := \hat{\phi}_1 = \hat{\mu} = 0$			$\hat{F}_3 := \hat{\phi}_2 = \hat{\mu} = 0$		
$T$	1%	5%	10%	1%	5%	10%	1%	5%	10%
50	9.31	6.73	5.61	7.02	5.13	4.31	7.06	4.86	3.94
100	8.73	6.49	5.47	6.50	4.88	4.16	6.70	4.71	3.86
250	8.43	6.34	5.39	6.22	4.75	4.07	6.52	4.63	3.81
500	8.34	6.30	5.36	6.15	4.71	4.05	6.47	4.61	3.79
$\infty$	8.27	6.25	5.34	6.09	4.68	4.03	6.43	4.59	3.78

# Les valeurs critiques de Dickey-Fuller

- Dickey-Fuller proposent des valeurs critiques tabulées pour

$$\hat{\tau}_1, \hat{\tau}_2, \hat{\tau}_3$$

Note 1 Les distributions limites ressemblent toutes à l'Eq. 6

Note 2 Elles diffèrent légèrement selon la présence de  $\mu \neq 0$  et/ou  $\delta \neq 0$ ,

$T$	M1 : $\hat{\tau}_1$			M2 : $\hat{\tau}_2$			M3 : $\hat{\tau}_3$		
	1%	5%	10%	1%	5%	10%	1%	5%	10%
50	-4.15	-3.50	-3.18	-3.58	-2.93	-2.60	-2.62	-1.95	-1.61
100	-4.04	-3.45	-3.15	-3.51	-2.89	-2.58	-2.6	-1.95	-1.61
250	-3.99	-3.43	-3.13	-3.46	-2.88	-2.57	-2.58	-1.95	-1.62
500	-3.98	-3.42	-3.13	-3.44	-2.87	-2.57	-2.58	-1.95	-1.62
$\infty$	-3.96	-3.41	-3.12	-3.43	-2.86	-2.57	-2.58	-1.95	-1.62

## Le test de Dickey-Fuller Augmenté : le modèle AR( $p$ )

- Le test précédent se limite au modèle AR(1) mais qu'en est-il si

$$X_t = \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + \varepsilon_t$$

car déterminer et tester les  $p$  racines est un tâche complexe ?

- Pour simplifier le problème introduisons  $\rho \equiv \phi_1 + \dots + \phi_p$  et

$$\xi_j \equiv -(\phi_{j+1} + \phi_{j+2} + \dots + \phi_p)$$

pour  $j = 1, 2, \dots, p-1$  de sorte qu'on puisse écrire

$$\begin{aligned} \varepsilon_t &= \left( (1 - \rho L) - (\xi_1 L + \xi_2 L^2 + \dots + \xi_{p-1} L^{p-1})(1 - L) \right) X_t \\ &= X_t - \rho X_{t-1} - \xi_1 \Delta X_{t-1} - \xi_2 \Delta X_{t-2} - \dots - \xi_{p-1} \Delta X_{t-p+1} \end{aligned}$$

et  $X_t = \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + \varepsilon_t$  de manière équivalente car

$$(1 - \rho L) - (\xi_1 L + \xi_2 L^2 + \dots + \xi_{p-1} L^{p-1})(1 - L) = 1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p$$

## Le test de Dickey-Fuller Augmenté : intuition

- La reformulation nous ramène vers un modèle proche du test DF

$$X_t = \rho X_{t-1} + \sum_{j=1}^p \xi_j \Delta X_{t-j} + \varepsilon_t$$

- Si une des racine de  $1 - \phi_1 z - \dots - \phi_p z^p$  est  $z = 1$ ,  $\rho = 1$  ou  $\phi = 0$

$$\Delta X_t = (\rho - 1)X_{t-1} + \sum_{j=1}^p \xi_j \Delta X_{t-j} + \varepsilon_t = \phi X_{t-1} + \sum_{j=1}^p \xi_j \Delta X_{t-j} + \varepsilon_t$$

- La formulation du test de racine unitaire reste alors

$$H_0 : \phi = 0 \text{ versus } H_1 : \phi < 0$$

- La distribution limite des OLS pour  $\phi$  reste celle du Th.3

⇒ la distribution de la statistique de “type” Student

$$\hat{\tau}_3 = \frac{\hat{\phi}_3}{\hat{\sigma}_\phi}$$

reste celle de l’Eq. 6 mais les  $\hat{\tau}_{\xi_j}$  ont une distribution standard

## La procédure de Dickey-Fuller Augmenté

- La procédure du test ADF est identique à celle du test DF
- Il est néanmoins crucial ici de déterminer  $p$  auparavant

0.a. Estimer le modèle

$$\Delta X_t = \mu + \delta t + \phi X_{t-1} + \sum_{j=1}^{pmax} \xi_j \Delta X_{t-j} + \varepsilon_t$$

- 0.b. Contrôler la présence d'auto-corrélation sérielle des résidus (test de Ljung-Box)
- 0.c. Si les résidus sont auto-corrélés on augmente  $pmax$
- 0.d. Si les résidus sont non auto-corrélés on teste avec  $p < pmax$
- 0.e. On garde le plus petit  $p$  qui satisfait 0.d. et on passe à l'étape 1.a.

Note L'utilisation des critères de sélection est aussi possible

## La procédure ADF en présence d'ARMA( $p, q$ )

- La procédure du test ADF reste identique
- ⇒ Elle repose la représentation AR( $\infty$ ) des ARMA( $p, q$ )
- Il faut chercher une **l'approximation** par un AR( $p < \infty$ )
- Dans ce context les critères AIC ou BIC sont à privilégier

Note La procédure de sélection de  $p$  est importante pour que

$$\hat{\tau}_1, \hat{\tau}_2, \hat{\tau}_3$$

possèdent les distributions (non-standard) attendues

## Le test de Phillips-Perron (PP) : intuition

- L'éradication de l'autocorrélation sérielle de  $\varepsilon_t$  peut être délicate
- ⇒ PP proposent une correction non-paramétrique basée sur

$$\lambda_n^2,$$

la variance de long terme de  $\varepsilon_t$

- L'estimateur proposé repose sur les autocovariances empiriques

$$\hat{\gamma}_n(j) = n^{-1} \sum_{t=j+1}^n \hat{\varepsilon}_t \hat{\varepsilon}_{t-j}$$

et s'exprime comme (l'estimateur de Newey-West)

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

- Une statistique corrigée de type Student en découle

$$\hat{\tau}_j^* = \sqrt{\frac{\hat{\gamma}_n(0)}{\hat{\lambda}_n^2} \frac{\hat{\rho} - 1}{\hat{\sigma}_{\hat{\rho}}}} - \frac{1}{2} (\hat{\lambda}_n^2 - \hat{\gamma}_n(0)) \frac{1}{\hat{\lambda}_n} \frac{n \hat{\sigma}_{\hat{\rho}}}{\hat{\sigma}_{\hat{\varepsilon}}}, \quad j = 1, 2, 3$$

## Le test de Phillips-Perron (PP) : implémentation

- Le test de PP s'applique de manière identique au test de DF
  - La correction dépend en revanche d'un paramètre de tuning  $q$
- ⇒ ce paramètre de troncature ne doit pas croître plus vite que

$$n^{1/4}$$

sous peine de non-convergence de  $\hat{\lambda}_n^2$  vers  $\lambda_n^2$

⇒ Sous cette condition, la correction est valide, et

$$\hat{\tau}_j^* \stackrel{d}{=} \hat{\tau}_j$$

Note Les distributions des statistiques de type Student sont identiques à DF et ADF pour chacun des 3 modèles

# Plan

- 1 Faits stylisés et rappels
- 2 DS versus TS
- 3 Intégration fractionnaire
- 4 ARMA Saisonniers
- 5 Théorie limite standard
- 6 Théorie limite non-standard
- 7 Tests de racine unitaire
- 8 Les régressions factices
- 9 Conclusion
- 10 Rappels de Probabilités

# 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 (4)

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  $\beta$

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

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

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

⇒ 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

$$\begin{pmatrix} \sigma_\eta^2 & 0 \\ 0 & \sigma_\nu^2 \end{pmatrix}^{-1/2} n^{-1/2} \sum_{t=1}^{[nr]} \begin{pmatrix} \eta_t \\ \nu_t \end{pmatrix} \xrightarrow{d} \begin{pmatrix} W_Y(r) \\ W_X(r) \end{pmatrix} \quad (9)$$

- 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 (9)
- 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  $n^{-1/2} \hat{t}_\beta$

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

- Phillips (1986) montre également qu'en présence d'une constante  $\alpha$  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

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

## Rappels : test de Durbin-Watson et $R^2$

- Le test DW est un test d'autocorrélation à l'ordre 1 pour

$$y_t = \beta x_t + u_t, \quad u_t = \varphi u_{t-1} + \varepsilon_t$$

avec  $|\varphi| < 1$  et  $\varepsilon_t \sim WN(0, \sigma_\varepsilon^2)$

- Le test requiert également que  $x_t \neq y_{t-1}$  et sous

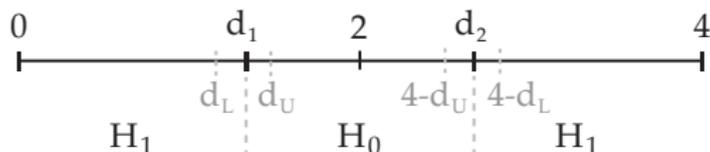
$$H_0 : \varphi = 0 \text{ versus } H_1 : \varphi \neq 0$$

voit sa statistique de test

$$DW = \frac{\sum_{t=2}^n (\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{t=2}^n \hat{u}_t^2} \approx 2(1 - \hat{\varphi}) \in [0, 4]$$

suivre une distribution non standard dont les valeurs critiques  $d_U$  et  $d_L$  sont tabulées et encadrent des seuils inconnus  $d_1$  et  $d_2$

$\Rightarrow$  Une lecture rapide pointerait vers un non rejet de  $H_0$  si  $DW \approx 2$



## Rappels : $R^2$

- Le coefficient de détermination est donné par

$$\widehat{R}^2 = 1 - \left( \sum_{t=1}^n y_t^2 \right)^{-1} \left( \sum_{t=1}^n \widehat{\varepsilon}_t^2 \right)$$

- La statistique  $\widehat{R}^2$  un estimateur de la quantité théorique

$$R^2 = 1 - \frac{\sigma_\varepsilon^2}{\sigma_y^2}$$

- Dans le cadre des ARMA( $p, q$ ), on ne connaît sa distribution limite

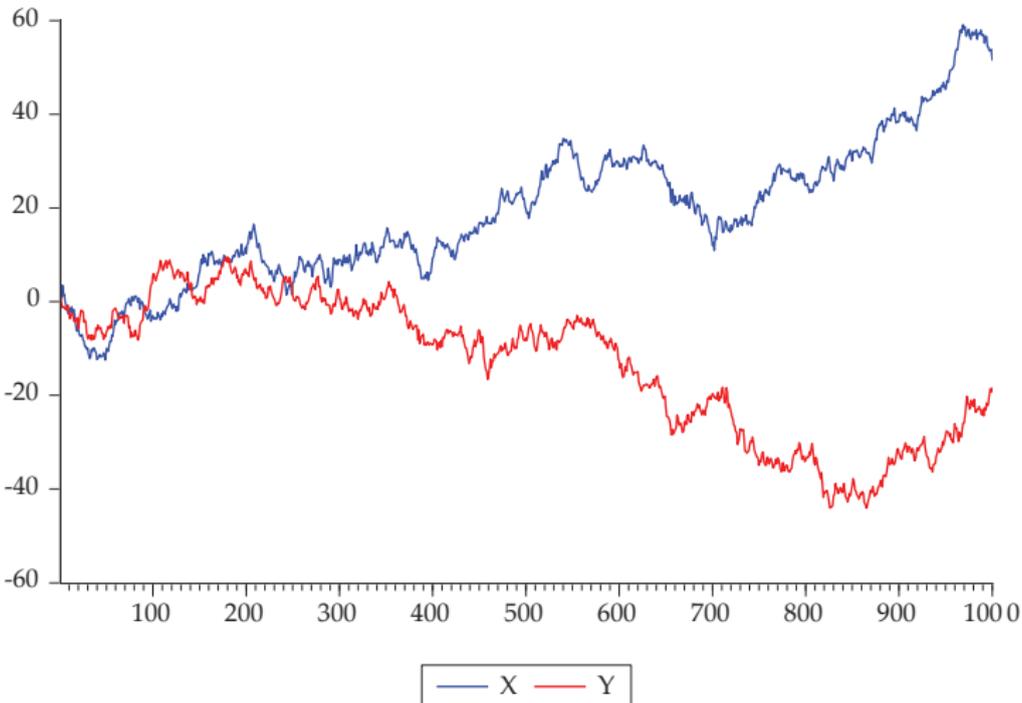
$$\sqrt{n}(\widehat{R}^2 - R^2) \sim \mathcal{N}\left(0, 4\sigma_\varepsilon^2(\sigma_y^2)^{-1} \sum_{i=1}^{\infty} \rho_i^2\right)$$

que pour  $q = 0$  et dans un cadre non-stationnaire, elle devient non-standard

- En présence de régression fallacieuse, elle est non-dégénérée
- ⇒  $\widehat{R}^2$  pointe vers des valeurs  $> 0$  et donc erronées
- ⇒  $DW$  pointe vers des valeurs anormalement basses

## Exemple : régression factice

- Soit  $X_t$  et  $Y_t$  deux marches aléatoires indépendantes



## Exemple : régression factice

- Estimons le modèle

$$X_t = \beta_0 + \beta_1 Y_t + \varepsilon_t$$

pour lequel on s'attend à trouver  $\hat{\beta}_1 = 0$

⇒ pourtant les résultats indiquent un relation significative

Included observations: 1000

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	9.557269	0.482904	19.79125	0.0000
Y	-0.700810	0.024364	-28.76363	0.0000
R-squared	0.453254	Mean dependent var		18.87462
Adjusted R-squared	0.452707	S.D. dependent var		15.30898
S.E. of regression	11.32547	Akaike info criterion		7.693984
Sum squared resid	128009.8	Schwarz criterion		7.703799
Log likelihood	-3844.992	Hannan-Quinn criter.		7.697714
F-statistic	827.3464	Du rbin-Watson stat		0.011903
Prob(F-statistic)	0.000000			

## Exemple : régression factice

- Estimons à présent le modèle

$$\Delta X_t = \beta_1 \Delta Y_t + u_t$$

pour lequel on s'attend à trouver  $\hat{\beta}_1 = 0$

⇒ à présent on constate qu'il n'y a aucune relation entre  $X_t$  et  $Y_t$

Included observations: 999 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.051278	0.031154	1.645935	0.1001
DY	0.035033	0.030722	1.140296	0.2544
R-squared	0.001302	Mean dependent var		0.050595
Adjusted R-squared	0.000301	S.D. dependent var		0.984653
S.E. of regression	0.984505	Akaike info criterion		2.808644
Sum squared resid	966.3420	Schwarz criterion		2.818467
Log likelihood	-1400.918	Hannan-Quinn criter.		2.812378
F-statistic	1.300276	Durbin-Watson stat		2.072721
Prob(F-statistic)	0.254437			

# 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  $\delta_y + \delta_x > 1/2$

⇒ 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
- 2 DS versus TS
- 3 Intégration fractionnaire
- 4 ARMA Saisonniers
- 5 Théorie limite standard
- 6 Théorie limite non-standard
- 7 Tests de racine unitaire
- 8 Les régressions factices
- 9 Conclusion
- 10 Rappels de Probabilités

## 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  $\delta_x + \delta_y > 1/2$

# Plan

- 1 Faits stylisés et rappels
- 2 DS versus TS
- 3 Intégration fractionnaire
- 4 ARMA Saisonniers
- 5 Théorie limite standard
- 6 Théorie limite non-standard
- 7 Tests de racine unitaire
- 8 Les régressions factices
- 9 Conclusion
- 10 Rappels de Probabilités

# 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)

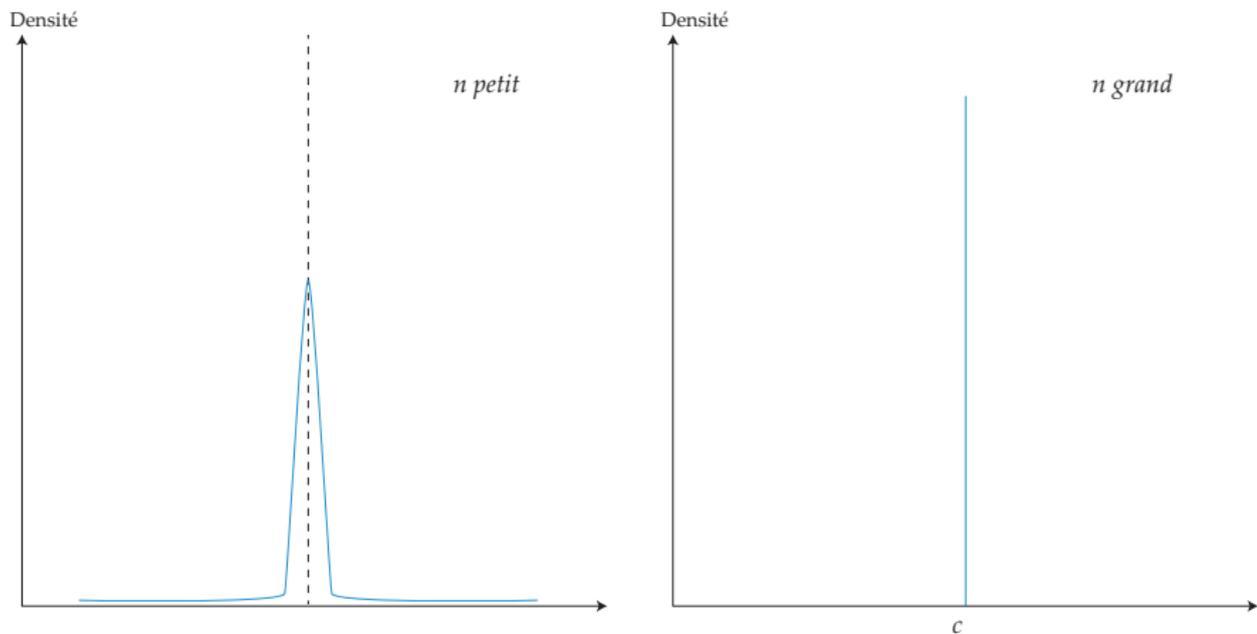
## Definition (7)

$x_n$  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

- **Implications** : quand  $n \rightarrow \infty$ ,  $X_n$  tend vers une constante

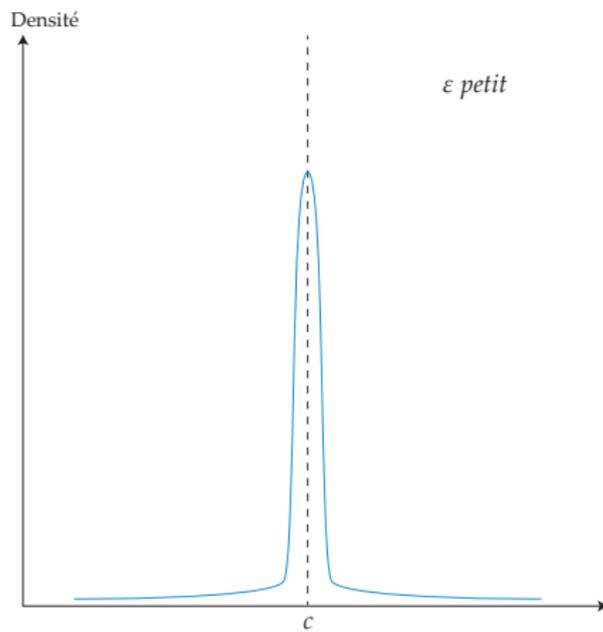
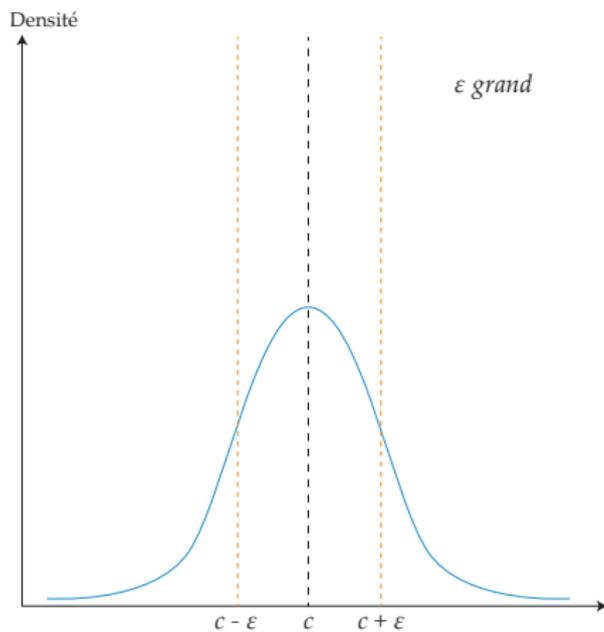
## Definition (8)

$X_n$  converge en probabilité vers une constante  $c$ , si pour toute valeur de  $\epsilon > 0$ ,

$$\lim_{n \rightarrow \infty} \Pr(|X_n - c| > \epsilon) = 0$$

- **Notation mathématique** :  $X_n \xrightarrow{p} c$  ou  $\text{plim } X_n = c$
- **Explications** : La convergence en probabilité n'est pas stricte comme la convergence presque sûre et on parle également de convergence au sens faible. Par conséquent,  $X_n$  converge asymptotiquement vers une quantité aléatoire dont la densité est très concentrée autour de  $c$

# Convergence en probabilité : Illustration



# Convergence en moyenne quadratique

- **Implications** : quand  $n \rightarrow \infty$ ,  $X_n$  tend vers une constante

## Definition (9)

$X_n$  converge en moyenne quadratique vers une constante  $c$ , si  $\mathbb{E}(|X_n|^2) < \infty$  et si pour toute valeur de  $\gamma > 0$ ,

$$\mathbb{E}(|X_n - c|^2) < \gamma$$

- **Notation mathématique** :  $X_n \xrightarrow{m.s.} c$
- **Explications** :  $X_n$  converge en moyenne quadratique si sa distribution est centrée sur  $c$ , i.e.  $\mathbb{E}(X_n) = c$  et si sa variance tend vers 0 asymptotiquement, impliquant une densité très concentrée autour de  $c$

## Convergence en loi

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

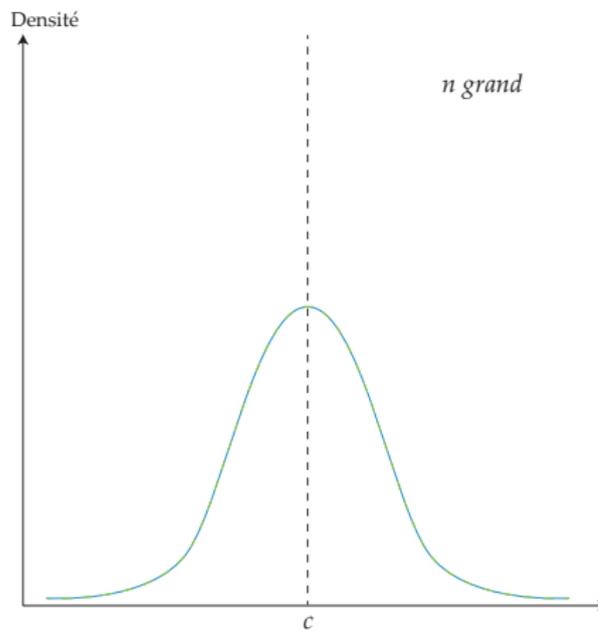
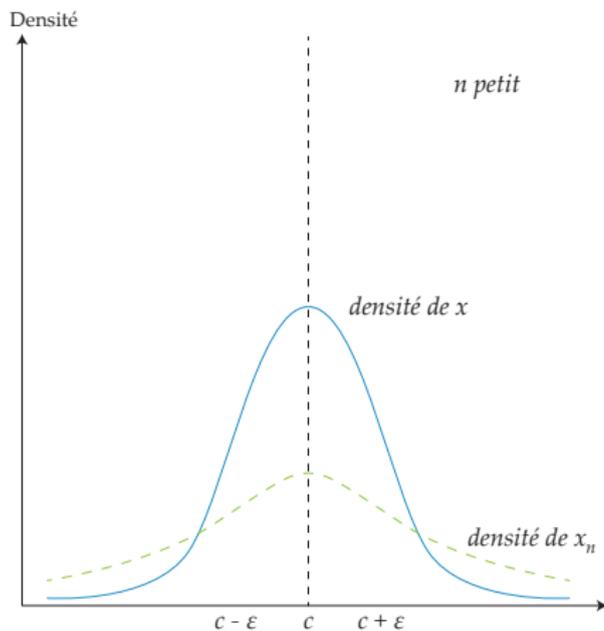
### Definition (10)

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 (5 : 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

$$\text{plim } \bar{X} = \lim_{n \rightarrow \infty} n^{-1} \sum_{t=1}^n \mathbb{E}(X_t) \text{ si } X_t \text{ indépendant } \forall t \quad (10)$$

$$= \mu \text{ si } X_t \sim \text{i.i.d.} \quad (11)$$

## Theorem (6 : 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  $X_t$*

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

# 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 (7 : 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

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

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

- Car :

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

# 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  $2\alpha - 1 = 0$

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

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

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

# Le théorème central limite

## Theorem (8 : 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 (9 : 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

# Théorème de Mann–Wald

## Theorem (10 : Continuous Mapping Theorem)

Soit  $X_t = X_1, \dots, X_n$  une séquence de variables aléatoires et  $g(\cdot)$  une fonction continue à valeur dans  $\mathbb{R}$  en  $X$ . Alors,

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

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

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

### ■ Exemples

$$\blacksquare (X_n, Y_n) \xrightarrow{p} (X, Y) \text{ implique } (X_n Y_n) \xrightarrow{p} XY$$

$$\blacksquare (X_n, Y_n) \xrightarrow{p} (X, Y) \text{ implique } (X_n/Y_n) \xrightarrow{p} X/Y \text{ si } Y \neq 0$$

# Théorème de Slutsky

## Theorem (11 : Théorème de Slutsky)

Soit  $X_t = X_1, \dots, X_n$  et  $Y_t = Y_1, \dots, Y_n$  deux séquences de variables aléatoires telles que  $X_t \xrightarrow{d} X$  et  $Y_t \xrightarrow{p} c \neq 0$ . Alors,

$$X_t + Y_t \xrightarrow{d} X + c, \quad X_t Y_t \xrightarrow{d} Xc, \quad \frac{X_t}{Y_t} \xrightarrow{d} \frac{X}{c}$$

- En supposant que  $X_t \xrightarrow{d} \mathcal{N}(m, \sigma^2)$  et que  $Y_t \xrightarrow{p} 2$ , le théorème nous apprend que

$$\frac{X_t}{Y_t} \xrightarrow{d} \mathcal{N}(m/2, \sigma^2/4)$$

- 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.