Chapitre I : Processus aléatoires stationnaires et processus ARMA

Econométrie des séries temporelles appliquée à la finance

Yannick LE PEN

- Econométrie des séries temporelles :
  - fondée sur les propriétés statistiques des séries
  - dépendance entre des observations à des dates différentes.
- On étudiera :
  - des processus en temps discret : observations quotidiennes, hebdomadaires, mensuelles, trimestrielles, annuelles.
  - des processus stationnaires et non stationnaires.
  - des processus univariés (une seule série) et multivariés (plusieurs séries).
- Il existe aussi des observations intra-daily (observations toutes les 5/10.. secondes, assimilables) à des observations en temps continue.
- On ne présentera pas ici les méthodes de modélisation des ces observations.

1. *Quel type de série ?*
   - Application de tests pour caractériser la composante déterministe et la composante stochastique.
2. *Comment modéliser ces séries ?*
   - Dépend des réponses aux tests
3. *Estimation et la spécification du modèle*
4. *Tests* de causalité, de cointégration, dans le cas multivarié.
5. *La prévision.*

- Un processus aléatoire (stochastique) est une suite ordonnée dans le temps de variables aléatoires $\{x_t\}$, $t \in \{0, 1, 2, 3, \ldots\}$.
- Dans la pratique, T observations : $\{x_1, x_2, \ldots.., x_T\}$.
- Une seule observation x_t pour chaque date
- Insuffisant pour estimer l'espérance $E(x_t) = \mu_t$ et ou la variance $V(x_t) = \sigma_t^2$ si elles dépendent de la date t.
- Nécessité d'imposer des contraintes sur le processus $\{x_t\}$.
- Si l'on impose les contraintes $E(x_t) = \mu_t = \mu$ et $V(x_t) = \sigma_t^2 = \sigma^2$, les T observations peuvent être utilisées pour estimer ces deux paramètres.

- Estimation de $E(x_t) = \mu_t = \mu$ par la moyenne empirique :

$$\bar{x}_T = \hat{\mu} = \frac{1}{T}\sum_{t=1}^{T} x_t$$

- Estimation de $V(x_t) = \sigma_t^2 = \sigma^2$ par la variance empirique :

$$\hat{\sigma}_T^2 = \frac{1}{T-1}\sum_{t=1}^{T} (x_t - \bar{x}_T)^2$$

ou

$$\tilde{\sigma}_T^2 = \frac{1}{T}\sum_{t=1}^{T} (x_t - \bar{x}_T)^2$$

Une mesure de la dépendance temporelle : l'autocovariance

- On veut mesurer la dépendance entre x_t et x_{t+h}.
- L'autocovariance $cov\left(x_t, x_{t+h}\right) = E\left(\left[x_t - E\left(x_t\right)\right]\left[x_{t+h} - E\left(x_{t+h}\right)\right]\right)$ au rang h mesure cette dépendance.
  - si $cov\left(x_t, x_{t+h}\right) = 0$, pas de dépendance entre x_t et x_{t+h}
  - si $cov\left(x_t, x_{t+h}\right) > 0$, il existe une relation positive entre x_t et x_{t+h}. Ils varient dans le même sens
  - si $cov\left(x_t, x_{t+h}\right) < 0$, il existe une relation négative entre x_t et x_{t+h}. Ils varient en sens opposés.
- Pour estimer l'autocovariance, on doit supposer qu'elle ne dépend pas de la date t mais seulement de l'écart h entre les dates t et $t+h$.

- Le processus aléatoire $\{x_t\}$, $t \in \{0, 1, 2, 3, \ldots\}$ est dit stationnaire du second ordre si et seulement si :
  1. L'espérance de x_t est la même pour toutes les dates :
  $$E(x_t) = \mu, \forall t \in \{0, 1, 2, 3, \ldots\}$$
  2. La variance de x_t est la même pour toutes les dates :
  $$V(x_t) = \sigma^2, \forall t \in \{0, 1, 2, 3, \ldots\}$$
  3. L'autocovariance est indépendante de la date t. Elle dépend seulement de la distance h entre deux observations :
  $$cov(x_t, x_{t+h}) = \gamma(h), \forall t \in \{0, 1, 2, 3, \ldots, \}, \forall h = 0, \pm 1, \pm 2, \pm 3, \ldots$$

Commentaires sur la notion de processus stationnaire

- permet de résoudre les problèmes d'estimation
- La fonction : $h \longmapsto \gamma(h) = cov(x_t, x_{t+h})$ est appelée fonction d'autocovariance.
- La fonction d'autocovariance est paire : $\forall h, \gamma(h) = \gamma(-h) \Rightarrow$ il suffit de connaître les autocorrélations pour $h \geq 0$.
- Par la suite, lorsque l'on emploiera les termes de "stationnaire" ou de "stationnarité", on désignera la stationnarité du second ordre.
- Toutes les séries temporelles ne sont pas stationnaires au second ordre.

Commentaires sur la notion de processus stationnaire

- *Notation*
- Notion de processus intégré d'ordre d.
- $d =$ nombre de fois où le processus doit être différencié pour se ramener à un processus stationnaire.
- Si x_t est stationnaire : $x_t \sim I(0)$
- Si est x_t DS alors $x_t \sim I(1)$ et $\Delta x_t \sim I(0)$

- Une mesure de la dépendance temporelle
- La fonction d'autocovariance est : $\gamma(h) = cov(x_t, x_{t+h})$
- $\gamma(h)$ dépend de l'unité de mesure de x_t
- La fonction d'autocorrélation au rang h est définie par le ratio : $\rho(h) = \frac{\gamma(h)}{\gamma(0)}$
- $\rho(h) \in [-1, +1] \Rightarrow$ l'autocorrélation ne dépend pas de l'unité de mesure de x_t
- $\rho(0) = 1$
- $\rho(h) = \rho(-h) \Rightarrow$ il suffit de connaître les autocorrélations pour $h \geq 0$.
- L'autocorrélogramme : représentation graphique des autocorrélations.

- Les relations entre x_t et x_{t+h} peuvent :
  1. transiter par les variables intermédiares $x_{t+1}, x_{t+2}, \cdots, x_{t+h-1}$,
  2. être directe.

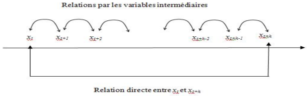


Figure – Relation directe et indirecte entre x_t et x_{t+h}

- le coefficient d'autocorrélation partielle r_h^h mesure la liaison linéaire *directe* entre x_t et x_{t+h} une fois retirés les liens transitant par les variables intermédiaires : $\{x_{t+1}, x_{t+2}, \ldots, x_{t+h-1}\}$.

Autre définition de l'autocorrélation partielle

- Si l'on régresse x_{t+h} sur $x_{t+h-1}, \dots, x_t$, on obtient :

$$x_{t+h} - \mu = r_h^1(x_{t+h-1} - \mu) + \dots\dots + r_h^h(x_t - \mu)$$

- Intérêt de l'autocorrélation partielle ?
- Permet de détecter des effets saisonniers.
- Pour des séries financières (5 jours par semaines), un effet ouverture des marchés le lundi peut être mis en évidence par de l'autocorrélation partielle au rang 6
- corrélation entre r_t et r_{t+6} indépendamment des jours intérmédiaires.

- $\gamma(h)$ est estimée par :

$$\hat{\gamma}(h) = \frac{1}{T-h} \sum_{t=1}^{T-h} (x_t - \bar{x}_T)(x_{t+h} - \bar{x}_T)$$

- Estimation de la fonction d'autocorrélation :

$$\hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}$$

- $\bar{x}_T = \frac{1}{T} \sum_{t=1}^{T} x_t$
- $\hat{\gamma}_0 = \hat{V}(x_t) = \frac{1}{T} \sum_{t=1}^{T} (x_t - \bar{x}_T)^2$

- Si $\rho(h) = 0$, on peut montrer que son estimation $\hat{\rho}(h) \in \left[\frac{-1.96}{\sqrt{T}}, \frac{1.96}{\sqrt{T}}\right]$ avec une probabilité de 95%.
- D'où, si $\hat{\rho}(h)$ sort des bornes de l'intervalle de confiance $\left[\frac{-1.96}{\sqrt{T}}, \frac{1.96}{\sqrt{T}}\right]$, on peut en conclure que $\rho(h) \neq 0$ avec une probabilité de 5% de se tromper (un risque de première espèce de 5%).

- Test vérifie la nullité simultanée des k premières autocorrélations :

$$\begin{cases} H_0 : \rho(1) = \rho(2) = ... = \rho(k) = 0 \\ H_1 : \exists i \in \{1, 2, ..., k\} \ \ t.q. \ \rho(i) \neq 0 \end{cases}$$

- Statistique du test de Ljung-Box :

$$Q_k^* = T\,(T+2)\sum_{j=1}^{k}\frac{1}{T-j}\hat{\rho}_j^2$$

- Sous H_0, $Q_k^* \to \chi^2(k)$
- On rejette H_0 si la statistique de test est supérieure au seuil critique choisi ou si la probabilité critique est inférieure à l'un des seuils habituellement choisis (1%, 5%, 10 %)

- Un estimateur de r_h^h est le coefficient de x_t de la régression par les MCO de x_{t+h} sur ses h dernières valeurs passées :

$$x_{t+h} = \hat{r}_1^h x_{t+h-1} + \ldots + \hat{r}_h^h x_t + \hat{u}_t$$

- On peut vérifier la nullité de r_h^h en utilisant la règle de décision suivante :
  - Si $r_h^h = 0$, $\hat{r}_h^h$ appartient à l'intervalle de confiance $\left[\frac{-1.96}{\sqrt{T}}, \frac{+1.96}{\sqrt{T}}\right]$ avec 95% de chance
  - Si $\hat{r}_h^h$ n'appartient pas à cet intervalle, on peut en conclure que $r_h^h \neq 0$ avec une probabilité de 5% de se tromper.

Exemples de séries stationnaires et non stationnaires

1. Etude graphique l'indice boursier S&P 500 et son taux de rendement,
2. Création de séries stationnaires ou non stationnaires par construction et voir leur aspect,
3. Comparaison de ces modèles de séries à l'indice S&P 500.

- période 03/01/2007 à 14/07/2017, données quotidiennes (5 jours par semaine, 2652 observations).

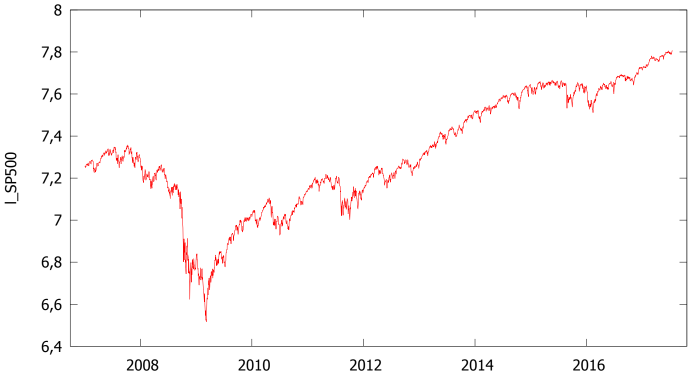
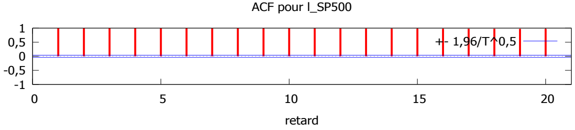
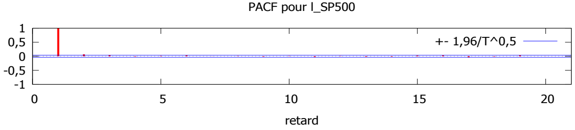


Figure – Indice S&P 500 en logarithme

- Série en logarithme népérien : on dit pour simplifier la série en log
- I_t = série en niveau (non transformée). On représente $i_t = ln(I_t)$
- Pourquoi le passage en log ?
  1. Contribue à lisser les fluctuations de la série,
  2. Interprétation de la série en différence comme un taux de variation (taux de rendement dans le cas d'une série financière).
- Cette série ne semble pas stationnaire au second ordre, car :
  1. Tendance à la baisse jusqu'en 2009, tendance à la hausse après 2009,
  2. Une persistence dans l'évolution de la série,
  3. Estimation de la moyenne et de la variance sur des sous-périodes.

Autocorrélations de l'indice S&P 500





1. $\hat{\rho}(h)$ proche de 1 (même pour h élevé) et significative,
2. *profil d'autocorrélation caractéristique d'une série non stationnaire.*
3. $\hat{r}_h^h$ proche de 1 pour $h = 1$.

Autocorrélations de l'indice S&P 500

Fonction d'auto-corrélation pour l_SP500
***, **, * indicate significance at the 1%, 5%, 10% leve
using standard error 1/T^0,5

RETARD	ACF		PACF		Q	[p. crit.]
1	0,9984	***	0,9984	***	2646,2994	[0,000]
2	0,9970	***	0,0694	***	5286,1259	[0,000]
3	0,9957	***	0,0403	**	7920,1071	[0,000]
4	0,9943	***	-0,0124		10547,9500	[0,000]
5	0,9930	***	0,0131		13169,9068	[0,000]
6	0,9918	***	0,0327	*	15786,5392	[0,000]
7	0,9906	***	-0,0025		18397,7152	[0,000]
8	0,9894	***	0,0096		21003,5894	[0,000]
9	0,9882	***	-0,0152		23603,9101	[0,000]
10	0,9869	***	0,0098		26198,8644	[0,000]
11	0,9857	***	-0,0267		28787,9913	[0,000]
12	0,9844	***	0,0083		31371,4948	[0,000]
13	0,9831	***	-0,0172		33949,1073	[0,000]
14	0,9817	***	-0,0128		36520,6536	[0,000]
15	0,9805	***	0,0238		39086,6082	[0,000]

- Taux de rendement de l'indice S&P 500 : différence première de l'indice en logarithme.

$$ln(I_t) - ln\left(I_{t-1}\right) = ln\left(\frac{I_t}{I_{t-1}}\right) \simeq r_t$$

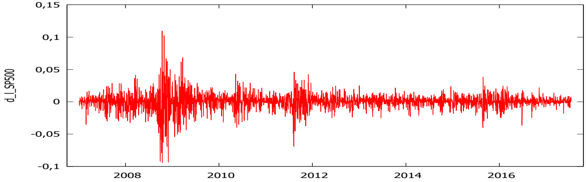
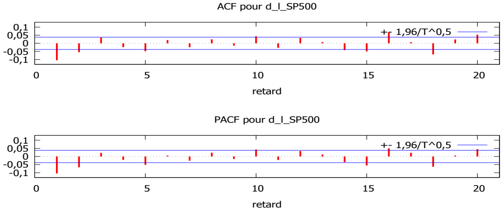


Figure – Taux de rendement de l'indice SP 500

- Pas de tendance à la hausse ou à la baisse dans le profil de la série,
- des pics de volatilité.

Autocorrélations du rendement du S&P 500



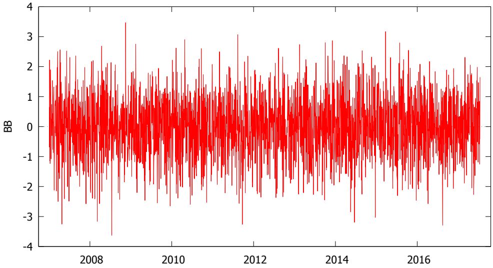
- $_r\hat{h}o(h)$ petite et décroit très rapidement.
- profil de corrélation est caractéristique d'une série stationnaire
- de même pour l'autocorrélation partielle.
- Le Q-test rejette toujours l'absence d'autocorrélation

- Les objectifs de ce paragraphe sont de :
- présenter des modèles types de séries temporelles,
- générer des échantillons aléatoires à partir de ces modèles,
- les comparer aux séries économiques vues précédemment,
- voir quels modèles pourraient reproduire les trajectoires de séries économiques observées.

Bruit blanc

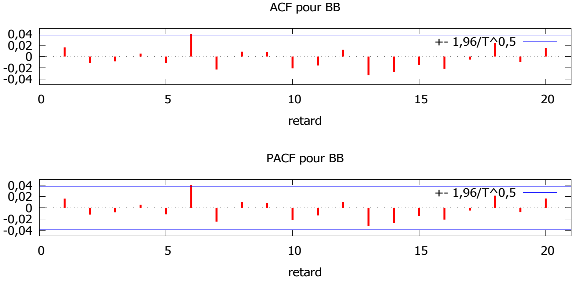
- Un bruit blanc est une suite de variables aléatoires $\epsilon_t, t \in \{0, 1, 2, ...\}$
  - de moyenne nulle : $E(\epsilon_t) = 0, \forall t \in \{0, 1, 2, ...\}$
  - de variance constante : $E(\epsilon_t^2) = V(\epsilon_t) = \sigma_\epsilon^2 = \gamma(0), \forall t \in \{0, 1, 2, ...\}$
  - non corrélées dans le temps : $\gamma(h) = cov(\epsilon_t, \epsilon_{t+h}) = 0, \forall t \in \{0, 1, 2, ...\}, \forall h \in \{1, 2, ...\}$,
- Notation : $\epsilon_t \sim BB(0, \sigma_\epsilon^2)$
- Un bruit blanc est stationnaire au second ordre par construction
- Un bruit blanc est un processus sans mémoire.
- Si, de plus, ϵ_t suit une loi normale, on a un bruit blanc gaussien, et on écrit : $\epsilon_t \sim NID(0, \sigma_\epsilon^2)$.

Exemple d'un processus bruit blanc



- L'indice S&P 500 ne ressemble pas à un bruit blanc
- Le taux de rendement du S&P 500 pourrait ressembler à un bruit blanc (à part les pics de volatilité)

Autocorrélogramme d'un bruit blanc



- $\hat{\rho}(h)$ presque jamais significative
- profil similaire de l'autocorrélation partielle
- Q-test ne rejette jamais l'hypothèse d'absence d'autocorrélation

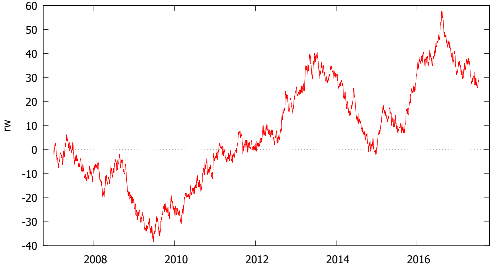
Marche aléatoire

- x_t tel que $x_t = x_{t-1} + \epsilon_t$ où $\epsilon_t \sim BB(0, \sigma_\epsilon^2)$ est une marche aléatoire.
- Une marche aléatoire n'est pas stationnaire du second ordre.
- La marche aléatoire x_t = somme cumulée des ϵ_t passés :

$$x_t = x_{t-1} + \epsilon_t = \epsilon_t + \epsilon_{t-1} + \ldots + \epsilon_1 = \sum_{i=1}^{t} \epsilon_i$$

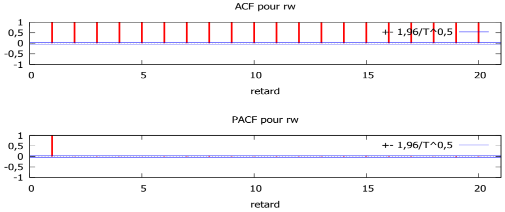
- L'impact d'un choc est permanent.
- $E(x_t) = 0$ et $V(x_t) = t\sigma_\epsilon^2$
- L'espérance de x_t est constante.
- La variance de x_t augmente avec le temps.
- Les conditions de stationnarité ne sont pas satisfaites

Marche aléatoire



- Cet exemple de marche aléatoire ressemble assez au profil du S&P 500.

Autocorrelogrammes d'une marche aléatoire



- $\hat{\rho}(h)$ très proche de 1 et décroit lentement.
- r_1^1 très proche de 1 et nulle pour les autres ordres.
- *Profil de corrélation est caractéristique d'une marche aléatoire.*
- Le Q-test rejette toujours l'hypothèse d'absence d'autocorrélation.

Marche aléatoire avec constante

- On ajoute une constante à la marche aléatoire :
  $x_t = c + x_{t-1} + \epsilon_t \Rightarrow \Delta x_t = c + \epsilon_t$.
- La constante c introduit une tendance déterministe dans l'expression de x_t :

$$
\begin{aligned}
x_1 &= x_0 + c + \epsilon_1 \\
x_2 &= x_1 + c + \epsilon_2 = x_0 + 2c + \epsilon_2 + \epsilon_1 \\
&\vdots \\
x_t &= x_0 + c \times t + \sum_{i=1}^{t} \epsilon_i
\end{aligned}
$$

Conclusions

- l'autocorrélogramme de l'indice S&P 500 est très proche de celui d'une marche aléatoire, c'est-àdire d'une série non-stationnaire
- l'autocorrélogramme du taux de rendement de l'indice S&P 500 est proche de celui d'un bruit blanc c'est à dire de celui d'un processus stationnaire.
- en différenciant l'indice S&P 500, on s'est ramené à une série stationnaire.

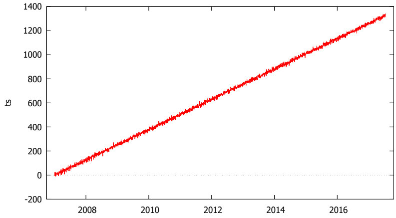
Comment se ramener à une série stationnaire à partir d'une marche aléatoire ?

- x_t tel que $x_t = x_{t-1} + \epsilon_t$, où $\epsilon_t \sim BB(0, \sigma_\epsilon^2)$
- la différence première de x_t est un processus stationnaire du second ordre : $\Delta x_t = x_t - x_{t-1} = \epsilon_t$
- x_t est stationnaire en différence
- x_t est un processus DS ("DS" pour "difference stationary".)

Processus stationnaire autour d'une tendance déterministe

- $x_t = \alpha + \beta t + \epsilon_t$ avec $\epsilon_t \sim BB\left(0, \sigma_\epsilon^2\right)$
- α et β sont des paramètres et *t=0,1,2,...* est la tendance déterministe
- La non-stationnarité de x_t provient de la présence de la composante déterministe.
- $E(x_t) = \alpha + \beta t \Rightarrow$ l'espérance de x_t dépend du temps.
- $V(x_t) = E[(x_t - E(x_t))^2] = E(\epsilon_t)^2 = \sigma_\epsilon^2 \Rightarrow$ la variance de x_t ne dépend pas de t
- $cov(x_t, x_{t+h}) = E[(x_t - E(x_t)(x_{t+h} - E(x_{t+h})] = E[\epsilon_t \epsilon_{t+h}] = 0, \forall h \neq 0 \Rightarrow$ l'autocorrélation est toujours nulle pour $h \neq 0$.

Processus stationnaire autour d'une tendance déterministe



- $\bar{x}_t \equiv x_t - (\alpha + \beta t) = \epsilon_t$ est stationnaire du second ordre, par définition de ϵ_t.
- x_t est *stationnaire autour d'un trend* ou encore *TS* pour "trend stationary".

3 modèles de l' économétrie des séries temporelles

1. le modèle autorégressif : AR
2. le modèle moyenne mobile : MA (pour moving average)
3. le modèle autorégressif-moyenne mobile (ARMA) qui rassemble les deux modèles précédents.

- L'opérateur de décalage L (ou B) permet de décaler la date d'une unité de temps :

$$Lx_t = x_{t-1}$$

- $L^n x_t = x_{t-n}$.
- $L^0 = 1$ d'où $L^0 x_t = x_t$.
- $Lc = c$ où c est une constante.
- L est un opérateur linéaire :
$(\sum_{i=0}^{n} a_i L^i)x_t = \sum_{i=0}^{n} a_i L^i x_t = \sum_{i=0}^{n} a_i x_{t-i}$.

Une série temporelle x_t suit un processus autorégressif d'ordre p si x_t dépend de :

1. Ses p valeurs passées : $x_{t-1}, x_{t-2}, x_{t-3}, \ldots, x_{t-p}$
2. Un bruit blanc ϵ_t de moyenne nulle et de variance σ_ϵ^2.
3. Le bruit blanc ϵ_t est la variation de x entre t et t-1 indépendante de ses valeurs passées.

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \cdots + \phi_p x_{t-p} + \epsilon_t = \sum_{i=1}^{p} \phi_i x_{t-i} + \epsilon_t$$

avec :

- $\{\phi_1, ..., \phi_p\}$ les p paramètres du modèle,
- $\epsilon_t, t \in \{0, 1, 2, ...\}$ un bruit blanc de moyenne nulle et de variance σ_ϵ^2.

- En utilisant l'opérateur de décalage L, on peut réécrire le modèle comme :

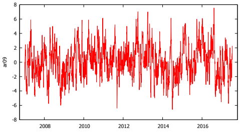
$$
\begin{aligned}
x_t - \sum_{i=1}^{p} \phi_i x_{t-i} &= \epsilon_t, \\
\Leftrightarrow (1 - \sum_{i=1}^{p} \phi_i L^i) x_t &= \epsilon_t, \\
\Leftrightarrow A(L) x_t &= \epsilon_t,
\end{aligned}
$$

- avec $A(L) = 1 - \sum_{i=1}^{p} \phi_i L^i$ et $A(0) = 1$.
- *Notation :* $x_t \sim AR(p)$

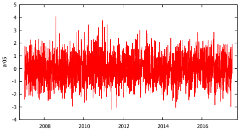
- La question qui se pose ici est de savoir si une série générée par un modèle *AR(p)* est stationnaire.
- L'équation $x_t = \sum_{i=1}^{p} \phi_i x_{t-i} + \epsilon_t$ décrit la dynamique de x_t
- Elle ne définit pas explicitement x_t en fonction de ϵ_t
- On n'est pas sur qu'elle définisse toujours un processus stationnaire au second ordre
- La stationnarité va dépendre des valeurs des paramètres $\phi_1, \phi_2, .., \phi_p$.
- A quelle condition sur les paramètres $\phi_1, \phi_2, .., \phi_p$ un processus autoregressif x_t est-il stationnaire ?
  1. Etude d'un processus AR(1)
  2. Généralisation à un AR(p)

- $x_t = \phi_1 x_{t-1} + \epsilon_t$
- La stationnarité de x_t dépend du paramètre ϕ_1.
- 3 cas :
  1. $\mid \phi_1 \mid < 1 \Rightarrow x_t$ est stationnaire
  2. $\mid \phi_1 \mid = 1 \Rightarrow x_t$ n'est pas stationnaire
  3. $\mid \phi_1 \mid > 1 \Rightarrow x_t$ n'est pas stationnaire

- Cas 1 : $\mid \phi_1 \mid < 1 \Rightarrow x_t$ est stationnaire.
- On génère deux processus AR(1) en prenant les valeurs $\phi_1 = 0.9$ et $\phi_1 = 0.5$. Des processus de ce type sont représentés ci-dessous :



(a) $\phi_1 = 0.9$



(b) $\phi_1 = 0.5$

- plus ϕ_1 est proche de 0, plus on se rapproche de l'aspect d'un bruit blanc
- plus ϕ_1 est proche de 1, plus on observe de la dépendance dans le profil de la série

Ecriture récursive d'un AR(1)

- Ecrire x_t en fonction des valeurs passées de ϵ_t : $x_0 = 0$

$$\begin{aligned}
x_0 &= 0,\ \textit{hypothèse simplificatrice} \\
x_1 &= \phi_1 x_0 + \epsilon_1 = \epsilon_1 \\
x_2 &= \phi_1 x_1 + \epsilon_2 = \phi_1 \epsilon_1 + \epsilon_2 \\
&\vdots \\
x_t &= \phi_1 x_{t-1} + \epsilon_t = \phi_1^{t-1}\epsilon_1 + \phi_1^{t-2}\epsilon_2 + \cdots + \epsilon_t
\end{aligned}$$

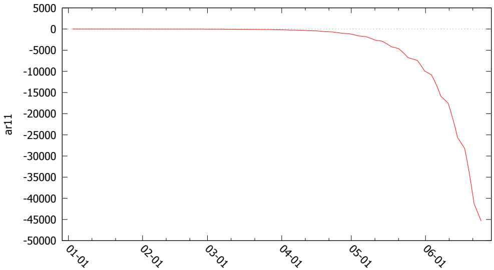
- l'impact des chocs passés ϵ_{t_i} sur x_t décroît avec le temps
- La série a tendance à revenir vers sa moyenne (ici $E(x_t) = 0$).

Condition de stationnarité d'un AR(1)

- Cas 2 : $\mid \phi_1 \mid = 1$.
- Quand $\phi_1 = 1$, on retrouve une marche aléatoire
- $x_t = x_{t-1} + \epsilon_t$. Cette série n'est pas stationnaire comme on l'a vu précedemment.
- Quand $\phi_1 = -1$, on a $x_t = -x_{t-1} + \epsilon_t$. Cette série n'est pas non plus stationnaire
- L'impact des chocs passés est permanent (1 ou -1).

Condition de stationnarité d'un AR(1)

- Cas 3 : $\mid \phi_1 \mid > 1$.
- Exemple : $\phi_1 = 1,1 \Rightarrow$ La série a une trajectoire explosive
- Peut être utilisé pour représenter des épisodes de bulle spéculative.



Reformulation de la condition de stationnarité d'un AR(1)

- Expression équivalente de la condition de stationnarité

$$\begin{aligned}
x_t &= \phi_1 x_{t-1} + \epsilon_t \\
\Rightarrow x_t - \phi_1 x_{t-1} &= \epsilon_t \\
\Rightarrow (1 - \phi_1 L)x_t &= \epsilon_t \\
\Rightarrow A(L)x_t &= \epsilon_t, \text{ avec } A(L) = 1 - \phi_1 L
\end{aligned}$$

- $\frac{1}{\phi_1}$ est racine de $A(L)$: $A(\frac{1}{\phi_1}) = 0$
- On a vu que x_t est stationnaire si $\mid \phi_1 \mid < 1$.
- On peut dire aussi que x_t est stationnaire si la racine $\frac{1}{\phi_1}$ de $A(L)$ est > 1.

Condition de stationnarité d'un AR(p)

- On considère un processus AR(p) quelconque :

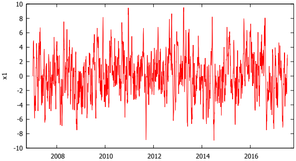
$$
\begin{aligned}
x_t &= \sum_{i=1}^{p} \phi_i x_{t-i} + \epsilon_t \\
\Rightarrow A(L)x_t &= \epsilon_t, \text{ avec } A(L) = 1 - \sum_{i=1}^{p} \phi_i x_{t-i}
\end{aligned}
$$

- Si les racines de $A(L)$ sont strictement supérieures à 1 (en valeur absolue), alors le processus x_t est stationnaire.
- Lors de l'estimation d'un processus AR, la plupart des logiciels reportent les racines ce qui permet de vérifier la stationnarité de la série considérée.

Condition de stationnarité d'un AR(p) : Exemples

- $x_{1,t}$ défini par le processus *AR(2)* :

$$\begin{aligned} x_{1,t} - 1,3x_{1,t-1} + 0,4x_{1,t-2} &= \epsilon_t \\ \Rightarrow (1 - 1,3L + 0.4L^2)x_{1,t} &= \epsilon_t \end{aligned}$$

- Racines de $A(L) = 2$ et 1,25 supérieures à 1 $\Rightarrow x_{1,t}$ est stationnaire.
- 

- $x_t = \phi_1 x_{t-1} + \epsilon_t$
- la condition de stationnarité est satisfaite :$\mid \phi_1 \mid < 1$
- Espérance de x_t : $E(x_t) = 0$
- Variance de x_t : $V(x_t) = \frac{\sigma_\epsilon^2}{1-\phi_1^2}$
- Fonction d'autocovariance de x_t :$\gamma(h) = \frac{\phi_1^h \sigma_\epsilon^2}{1-\phi_1^2}$
- Fonction d'autocorrélation de x_t : $\rho(h) = \phi_1^h$
- Fonction d'autocorrélation partielle :

$$\begin{cases} r_1^1 = \phi_1 \\ r_k^k = 0, \forall k > 1 \end{cases}$$

- Cette propriété est caractéristique des processus AR(1).

Propriété caractéristique d'un processus $AR(p)$

- La fonction d'autocorrélation partielle d'un processus $AR(p)$ s'annule à partir du rang $p+1$.
- C'est une propriété caractéristique d'un AR(p).
- Elle permet de savoir si une série doit être modélisée par un AR(p).

- x_t suit un processus moyenne mobile d'ordre q s'il s'écrit :

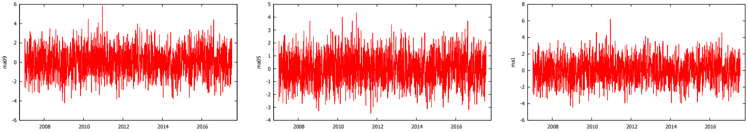
$$x_t = \epsilon_t + \theta_1 \epsilon_{t-1} + ... + \theta_q \epsilon_{t-q}, \forall t = 0, 1, 2, ...$$

- $\epsilon_t \sim BB(0, \sigma_\epsilon^2)$
- $\theta_1, \theta_2, ..., \theta_q$ sont des paramètres.
- Ecriture équivalente :

$$\begin{aligned} x_t &= \epsilon_t + \theta_1 \epsilon_{t-1} + ... + \theta_q \epsilon_{t-q} \\ \Rightarrow x_t &= (1 + \theta_1 L + ... + \theta_q L^q)\epsilon_t \\ \Rightarrow x_t &= B(L)\epsilon_t \end{aligned}$$

- Un processus $MA(q)$ est défini comme la somme des valeurs passées et présente d'un bruit blanc.
- Il est donc stationnaire par construction.

Exemples de processus MA(1)



(a) $\theta_1 = 0.9$ (b) $\theta_1 = 0.5$ (c) $\theta_1 = 1$

Figure – Trois processus MA(1)

- Les trois graphiques montrent des séries qui semblent stationnaires
- il y a relativement peu de différences évidentes entre ces séries
- On peut même remarquer qu'elles ressemblent au processus AR(1) avec $\phi_1 = 0.5$.

- $x_t = \epsilon_t + \theta_1 \epsilon_{t-1}$ où $\epsilon_t \sim BB(0, \sigma_\epsilon^2)$.
- Espérance de x_t : $E(x_t) = E(\epsilon_t + \theta_1 \epsilon_{t-1}) = 0$
- Variance de x_t : $V(x_t) = \sigma_\epsilon^2 (1 + \theta_1^2)$
- Fonction d'autocovariance :
$$\gamma(h) = E(x_t x_{t+h}) = \begin{cases} \gamma(1) = \theta_1 \sigma_\epsilon^2 \\ \gamma(h) = 0, \forall h > 1 \end{cases}$$
- Fonction d'autocorrélation : $\rho(h) = \frac{\gamma(h)}{\gamma(0)} \begin{cases} \rho(1) = \frac{\theta_1}{1+\theta_1^2} \\ \rho(h) = 0, \forall h > 1 \end{cases}$
- la fonction d'autocorrélation est caractéristique d'un processus $MA(1)$

- Pour tout processus $MA(q)$ $x_t = \epsilon_t + \theta_1\epsilon_{t-1} + ... + \theta_q\epsilon_{t-q}$
- Espérance de x_t : $E(x_t) = 0$
- Variance de x_t : $V(x_t) = (1 + \theta_1^2 + ... + \theta_q^2)\sigma_\epsilon^2$
- Fonction d'autocovariance de x_t :
$$\begin{cases} \gamma(h) = 0, \forall h > q \\ \gamma(h) = \sigma_\epsilon^2(\theta_h + \theta_{h+1}\theta_1 + ... + \theta_q\theta_{q-h}), \forall h \in \{1, \cdots, q\} \end{cases}$$
- Fonction d'autocorrélation de x_t :
$$\begin{cases} \rho(h) = 0, \forall h > q \\ \rho(h) = \frac{\theta_h + \theta_{h+1}\theta_1 + ... + \theta_q\theta_{q-h}}{1 + \theta_1^2 + ... + \theta_q^2}, si 1 \leq h \leq q \end{cases}$$

- Les processus $AR(p)$ et $MA(q)$ se distinguent par les propriétés sur les fonctions d'autocorrélation :
- $x_t \sim MA(q) \Rightarrow \rho(h) = 0, \forall h > q$
- $x_t \sim AR(p) \Rightarrow r_k^k = 0, \forall h > p$
- Peut-être utilisé pour différencier un processus AR(p) et MA(q).

- Un modèle *ARMA(p,q)* contient une composante autoregressive *AR(p)* et une composante moyenne mobile *MA(q)*. Un processus *ARMA(p,q)* s'écrit :

$$\begin{aligned} x_t - \phi_1 x_{t-1} - .... - \phi_p x_{t-p} &= \epsilon_t + \theta_1 \epsilon_{t-1} + ... + \theta_q \epsilon_{t-q} \\ \Rightarrow A(L) x_t &= B(L) \epsilon_t \end{aligned}$$

- $\epsilon_t \sim BB\left(0, \sigma_\epsilon^2\right)$.

Conditions de validité d'un ARMA(p,q)

- Les conditions suivantes doivent être vérifiées :
- $A(L)$ a ses racines strictement supérieures à 1 en valeur absolue : x_t est stationnaire
- $\phi_p \neq 0$ et $\theta_q \neq 0$: il n'y a pas de variable retardée superflue
- Les polynômes $A(L)$ et $B(L)$ n'ont pas de racines communes.
- Les deux dernières conditions visent à avoir le modèle le plus parcimonieux en termes de paramètres.

- On génère une série selon le processus ARMA(2,1) : $x_{3,t} = 0,9x_{3,t-1} - 0,2x_{3,t-2} + \epsilon_t - 0,8\epsilon_{t-1}$. Les racines du polynôme $A(L) = 1 - 0,9L + 0,2L^2$ sont égales à 2 et 2, 5. La série x_{3t} est donc stationnaire.

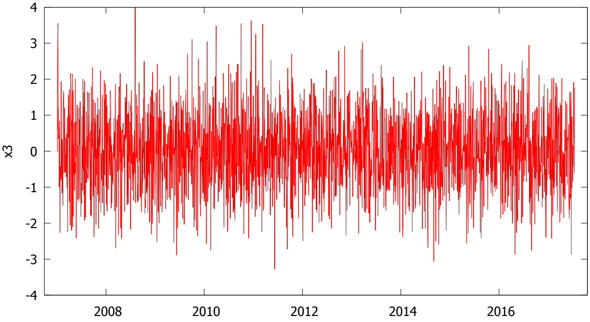
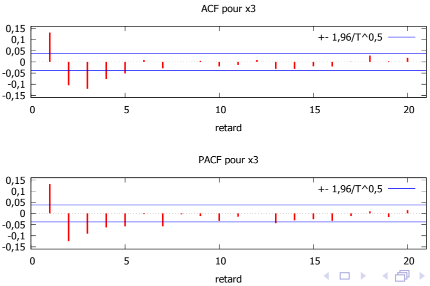


Figure – Représentation d'un processus stationnaire ARMA(2,1)

- Pour *ARMA(p,q)* stationnaire, la fonction d'autocorrélation $\rho(h)$ et la fonction d'autocorrélation partielle r_h^h décroissent vers 0 à mesure que h augmente.
- Pas d'annulation de $\rho(h)$ et r_h^h à partir des rangs p ou q.

Figure – Autocorrélogramme du processus *ARMA(1,1)*



- ARMA(p,q) avec une constante c :

$$x_t - \phi_1 x_{t-1} - .... - \phi_p x_{t-p} = c + \epsilon_t + \theta_1 \epsilon_{t-1} + ... + \theta_q \epsilon_{t-q}$$

- la constante c permet d'obtenir $E(x_t) \neq 0$

$$E(x_t) = \frac{c}{1 - \phi_1 - ..... - \phi_p} = \mu$$

- Si $c = 0$, $E(x_t) = 0$
- S'il y a une composante AR, $c \neq E(x_t)$
- Pour un processus $MA(q)$: $E(x_t) = c$

Construction d'une série de moyenne nulle

- On peut toujours se ramener à un processus y_t de moyenne nulle (centré)

$$y_t = x_t - E(x_t) = x_t - \frac{\theta}{1 - \phi_1 - ..... - \phi_p}$$

- y_t admet la même représentation ARMA(p,q) que x_t mais sans la constante c :

$$y_t - \phi_1 y_{t-1} - .... - \phi_p y_{t-p} = \epsilon_t + \theta_1 \epsilon_{t-1} + ... + \theta_q \epsilon_{t-q}$$

- Soustraire la moyenne ne modifie pas la dynamique de la série.