

Chapitre 2 : Stationnarité et autocorrélation simple et partielle

Avant de commencer l'analyse d'une série chronologique il faut étudier ses caractéristiques stochastiques. Si l'espérance et la variance de la série se trouve inchangé dans le temps la série est stationnaire, dans le cas contraire la série est non stationnaire. D'une façon formalisé la série est stationnaire si

$E(Y_t) = E(Y_{t+h}) = \mu \forall t \text{ et } \forall h$ La moyenne est constante et indépendante du temps ;

$V(Y_t) \approx \infty \forall t$ La variance est infinie et indépendante du temps ;

$Cov(Y_t, Y_{t+h}) = E[(Y_t - \mu)(Y_{t+h} - \mu)] = \gamma_h$ La covariance est indépendante du temps ;

2.1 Processus Stationnaire

Un processus est dit fortement stationnaire si toutes ses caractéristiques stochastiques sont invariantes dans le temps.

Le Processus Bruit Blanc est un processus stationnaire constitué de variables qui sont indépendantes, normalement distribué (loi normale centré réduite) et qui ont la même variance.

Le processus Bruit Blanc gaussien : est un processus stationnaire constitué de variables indépendantes qui suivent une loi normale et qui ont une variance identique.

La marche au hasard est un cas particulier d'un processus stochastique pour lequel la valeur prise par X à la date T est régie par l'équation suivante : $X_t = X_{t-1} + \varepsilon_t$ ou ε_t est une variable aléatoire qui a les mêmes caractéristiques que X

2.1.1 La fonction d'autocorrélation simple et d'autocorrélation partielle

La fonction d'autocorrélation

La fonction d'autocovariance d'un processus Y est donnée

$$k \in Z \rightarrow \gamma_k = E(Y_t Y_{t-k}) \in R$$

La fonction est estimée par :

$$R_k = \frac{1}{n-k} \sum_{t=k+1}^n Y_t Y_{t-k}$$

Avec

$$R_0 = \frac{1}{n} \sum_{t=k}^n Y_t^2 = V(Y_t)$$

La fonction d'autocorrélation de Y :

$$k \in Z \rightarrow \rho_k \in \#[-1, +1]$$

$$\rho_k = \frac{\gamma_k}{\gamma_0}$$

La variance de l'autocorrélation est donnée :

$$V(\rho_k) = \frac{1}{n} \sum_{i=k}^K \rho_i^2 \text{ Avec } k < K$$

ρ_i est considéré comme le coefficient d'autocorrélation théorique et r_i le coefficient de corrélation estimé, la variance de l'autocorrélation est donnée :

$$V(r_i) = \frac{1}{n} \left(1 + 2 \sum_{i=k}^K r_i^2 \right)$$

La fonction d'autocorrélation (ACF) est une fonction noté r qui mesure la correction de la série avec elle-même décalé de k période.

Elle est formulée comme suit :

$$r = \frac{cov(Y_t, Y_{t-k})}{\delta_{Y_t} \delta_{Y_{t-k}}} = \frac{\sum_{i=1}^n (Y_t - \bar{Y})(Y_{t-k} - \bar{Y})}{\sqrt{\sum (Y_t - \bar{Y})^2} * \sqrt{\sum (Y_{t-k} - \bar{Y})^2}}$$

n : représente le nombre d'observation ; $\bar{Y}$ représente la moyenne de la série calculée sur $n-k$ période.

Après transformation, on utilise la formule d'autocorrélation d'échantillonnage

$$\hat{r} = \frac{\sum_{i=1}^n (Y_t - \bar{Y})(Y_{t-k} - \bar{Y})}{\sum (Y_t - \bar{Y})^2}$$

$\bar{Y}$ représente la moyenne de la série calculée sur la période n

La fonction d'autocorrélation partielle

La fonction d'autocorrélation partielle (PACF) : la fonction d'autocorrélation partielle de X à distance h est la corrélation entre X_0 et X_h lorsque les valeurs intermédiaires sont fixées ($X_1, X_2, \dots, X_h$).

Soit la matrice ρ_k symétrique formée des $k - 1$ premières d'autocorrélation du processus Y_t .

$$\rho_k = \begin{pmatrix} 1 & \rho_1 & \rho_2 & \dots & \rho_{k-1} \\ \rho_1 & 1 & \rho_1 & \dots & \rho_{k-2} \\ \rho_2 & \rho_1 & 1 & \dots & \rho_{k-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \dots & 1 \end{pmatrix}$$

La fonction d'autocorrélation est la succession $\rho_{Kk} = \frac{\hat{\rho}_K}{\rho_k}$ avec

$|\hat{\rho}_K|$ Le déterminant de la matrice ρ_k dans laquelle on a remplacé la dernière colonne par le vecteur $(\rho_1, \dots, \rho_k)$

$$\hat{\rho}_K = \begin{pmatrix} 1 & \rho_1 & \rho_2 & \dots & \rho_1 \\ \rho_1 & 1 & \rho_1 & \dots & \rho_2 \\ \rho_2 & \rho_1 & 1 & \dots & \rho_3 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \dots & \rho_k \end{pmatrix}$$

Pour les grands échantillon les coefficients ρ_k du processus AR(P) sont asymptotiquement distribués selon une loi normale d'espérance nulle et de variance $V(\rho_k) = \frac{1}{n}$ pour $k > \rho$ (Quenouille, 1949). En effet, il est possible de construire un test d'hypothèse pour ρ_k .

Sous H_0 $\rho_k = 0$

Nous calculons $t^* = \frac{|\rho_k|}{\sqrt{\frac{1}{n}}}$

Cette statistique est comparée avec la statistique tablé de la table de la loi normale au seuil de signification de 95%.

Code Python pour générer les coefficients de corrélation simple et partiel

```
from statsmodels.tsa.stattools import acf
import matplotlib.pyplot as plt

acf_values = acf(x, nlags=50)

plt.stem(acf_values)

plt.title("ACF avec statsmodels")

plt.xlabel("Lag")

plt.ylabel("Autocorrélation")

plt.show()
```

2.2 Test de Bruit Blanc et de stationnarité

Avant d'identifier les caractéristiques d'une série stochastique il faut étudier sa stationnarité. Cette étude s'effectue principalement à partir de l'étude de la fonction d'autocorrélation (ou la représentation graphique qui est appelé Corrélogramme).

Une série chronologique est stationnaire si elle ne dispose ni de tendance ni de saisonnalité.

Afin d'effectuer l'analyse de stationnarité, il faut identifier les termes ρ_k qui sont significativement différent de 0.

2.2.1 Test de Box-Pierce et de Ljung-Box

Ce test permet d'identifier les processus Bruit Blanc. Les hypothèses du test sont données comme suit :

- $H_0 : \rho_1 = \rho_2 = \dots \rho_i = 0$
- $H_1 : \text{Il existe au moins un coefficient } \rho_i \neq 0$

Pour effectuer le test on calcule la statistique de Box-Pierce noté Q_{stat}

$$Q_{stat} = \frac{\sum_{i=1}^n (Y_t - \bar{Y})(Y_{t-k} - \bar{Y})}{\sum (Y_t - \bar{Y})^2}$$

Où La statistique est calculée pour chaque retard h de la manière suivante :

$$Q - stat = n(n + 2) \sum_{k=1}^p \frac{r_k^2}{n - k}$$

Avec n : nombre d'observation, k : nombre de retard. r représente la valeur du terme du corrélogramme simple au retard p.

Nous distinguons deux situations :

Si $Q_{stat} > X_n^2$ On rejette H_0 et on accepte H_1 L'hypothèse de Bruit Blanc est rejeté au seuil de signification de 5%.

Si $Q_{stat} < X_n^2$ On accepte H_0 . En effet, l'hypothèse de Bruit Blanc est acceptée.

Code Python pour le test de Box Pierce et de Ljung Box

```
import numpy as np
import statsmodels.api as sm
from statsmodels.stats.diagnostic import acorr_ljungbox

# -----
# Exemple : génération d'un bruit blanc
# -----
np.random.seed(42)
x = np.random.randn(300) # bruit blanc

# -----
# Test de Box-Pierce & Ljung-Box
# -----
# nlags : nombre de décalages à tester
results = acorr_ljungbox(x, lags=[10], return_vec=True)

print("=== Test de Ljung-Box ===")
```

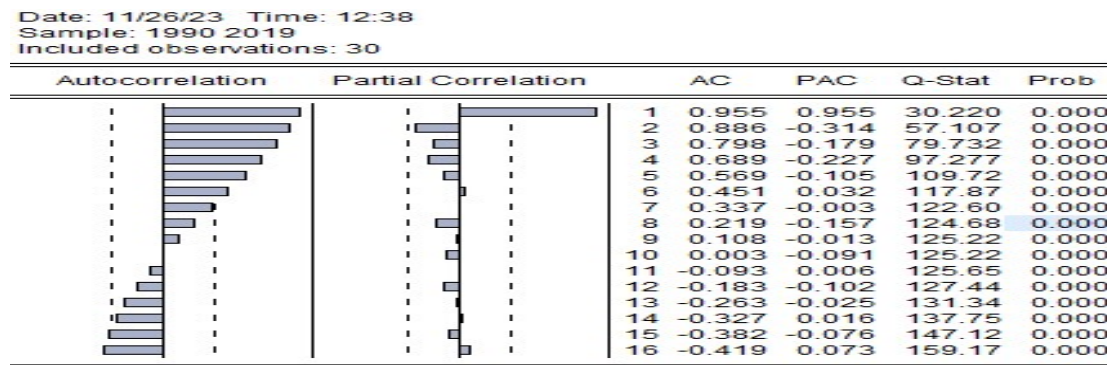
```
print(f"Statistic : {results.lb_stat[0]:.4f}")
print(f"p-value : {results.lb_pvalue[0]:.4f}")

print("\n=== Test de Box-Pierce ===")
print(f"Statistic : {results.bp_stat[0]:.4f}")
print(f"p-value : {results.bp_pvalue[0]:.4f}")
```

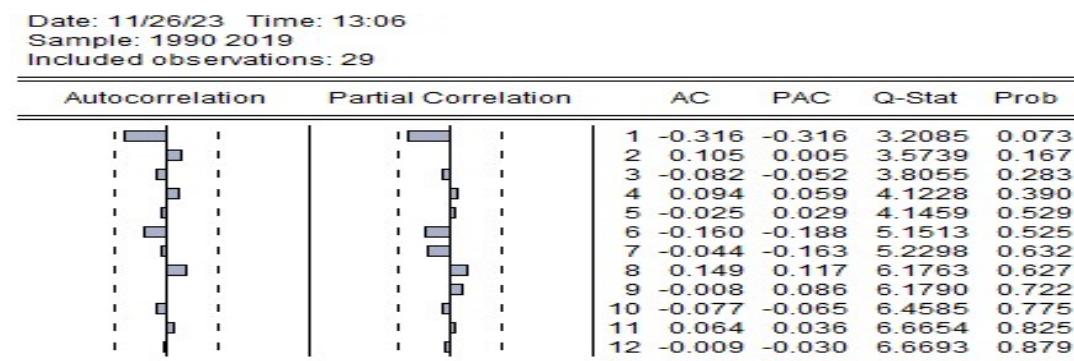
2.3 Analyse de la fonction d'autocorrélation

Lorsque nous analysons la fonction d'autocorrélation d'une série chronologique. La question qui se pose est celle de quels sont les termes ρ_i qui sont significativement différents de 0. Si par exemple aucun terme n'est significativement différent de 0 cela veut dire que la série ne dispose ni de tendance ni de saisonnalité donc elle est stationnaire.

Nous pouvons identifier le processus Bruit Blanc à partir du corrélogramme de la série, plus exactement le graphe d'autocorrélation.



Les premières valeurs de l'autocorrélation sont significativement différentes de 0. La série proposée n'est pas un Bruit Blanc.



L'ensemble des valeurs sont à l'intérieur de l'intervalle de confiance. L'ensemble des valeurs sont non significatives. La série proposée est un bruit blanc.

Série d'exercices n°2

Exercice n°1

L'évolution d'une série chronologique sur la période 1990-2009 est donnée comme suit :

Date	Y
1990	1,45
1991	1,41
1992	1,37
1993	1,34
1994	1,36
1995	1,37
1996	1,35
1997	1,33
1998	1,42
1999	4,41
2000	1,43
2001	1,46
2002	1,49
2003	1,51
2004	1,54
2005	1,58
2006	1,6
2007	1,65
2008	1,7
2009	1,73

1. Effectuer le test de Box-pierces et de Lyung Box.
2. Interpréter les résultats du test. Sachant que la statistique de Khi-deux $\chi^2_{20} = 10,85$

Le corrélogramme de la série est données comme suit :

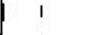
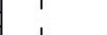
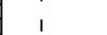
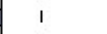
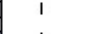
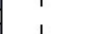
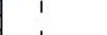
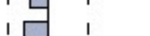
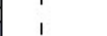
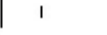
Date: 11/26/23 Time: 15:08
 Sample: 1990 2019
 included observations: 20

Autocorrelation			Partial Correlation			AC	PAC	Q-Stat	Prob	
						1	-0.060	-0.060	0.0832	0.773
						2	-0.087	-0.091	0.2666	0.875
						3	-0.075	-0.087	0.4107	0.938
						4	-0.070	-0.091	0.5456	0.969
						5	-0.067	-0.097	0.6783	0.984
						6	-0.070	-0.110	0.8321	0.991
						7	-0.060	-0.113	0.9522	0.996
						8	-0.040	-0.106	1.0113	0.998
						9	-0.014	-0.090	1.0192	0.999
						10	0.054	-0.022	1.1479	1.000
						11	0.009	-0.054	1.1523	1.000
						12	0.004	-0.052	1.1530	1.000

3. La série est-elle stationnaire ?

Exercice n°2

Sample: 1990 2019
Included observations: 30

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
		1	0.900	0.900	26.804	0.000
		2	0.814	0.024	49.538	0.000
		3	0.726	-0.057	68.278	0.000
		4	0.638	-0.054	83.284	0.000
		5	0.526	-0.175	93.913	0.000
		6	0.418	-0.072	100.91	0.000
		7	0.315	-0.046	105.06	0.000
		8	0.213	-0.071	107.05	0.000
		9	0.116	-0.044	107.66	0.000
		10	0.031	-0.013	107.71	0.000
		11	-0.017	0.124	107.72	0.000
		12	-0.097	-0.211	108.22	0.000
		13	-0.163	-0.037	109.72	0.000
		14	-0.224	-0.066	112.73	0.000
		15	-0.279	-0.094	117.71	0.000
		16	-0.326	-0.007	125.01	0.000

Soit le corrélogramme ci-dessus de l'indice CAC40.

1. Calculer la statistique de Lyang-Box pour le retard $h=4$.
2. Effectuer le test de Box-pierces et de Lyang-Box. $\chi_4^{0,95} = 9,48$

Exercice n°3

Soit les valeurs de la série Y suivants :

T	1	2	3	4	5	6	7	8	9	10
Y	10	3	-1	3	2	5	3	2	-1	3

1. Calculer les 03 premiers termes du corrélogramme simple.
2. Donner l'intervalle de confiance à 95% pour ces termes calculés et tester leur significativité.
3. Calculer les 03 premiers termes du corrélogramme partiel.
4. Donner l'intervalle de confiance pour ces termes au seuil de 95% et tester leur significativité.
5. Calculer la statistique de Ljung-Box pour $h=3$, la série est-elle un bruit-blanc ?

Solution de l'exercice n°1

4

5

	Y	Y_{t-1}	$Y - \bar{Y}$	$Y_{t-1} - \bar{Y}$	4*5	$(Y - \bar{Y})^2$
1	1,45		-0,023			0,000529
2	1,41	1,45	-0,063	-0,023	0,001449	0,003969
3	1,37	1,41	-0,103	-0,063	0,006489	0,010609
4	1,34	1,37	-0,133	-0,103	0,013699	0,017689
5	1,37	1,34	-0,103	-0,133	0,013699	0,010609
6	1,36	1,37	-0,113	-0,103	0,011639	0,012769
7	1,37	1,36	-0,103	-0,113	0,011639	0,010609
8	1,35	1,37	-0,123	-0,103	0,012669	0,015129
9	1,33	1,35	-0,143	-0,123	0,017589	0,020449
10	1,42	1,33	-0,053	-0,143	0,007579	0,002809
11	1,41	1,42	-0,063	-0,053	0,003339	0,003969
12	1,46	1,41	-0,013	-0,063	0,000819	0,000169
13	1,49	1,46	0,017	-0,013	-0,000221	0,000289
14	1,51	1,49	0,037	0,017	0,000629	0,001369
15	1,54	1,51	0,067	0,037	0,002479	0,004489
16	1,58	1,54	0,107	0,067	0,007169	0,011449
17	1,6	1,58	0,127	0,107	0,013589	0,016129
18	1,67	1,6	0,197	0,127	0,025019	0,038809
19	1,7	1,67	0,227	0,197	0,044719	0,051529
20	1,73	1,7	0,257	0,227	0,058339	0,066049
Somme					0,252331	0,29942

$$\bar{Y} = 1,473$$

1. le test de Box-pierces et de Lyung Box.

Ce test permet d'identifier les processus Bruit Blanc. Les hypothèses du test sont données comme suit :

- $H_0 : \rho_1 = \rho_2 = \dots \rho_i = 0$
- $H_1 : \text{Il existe au moins un coefficient } \rho_i \neq 0$

Pour effectuer le test on calcule la statistique de Box-pierce noté Q_{stat}

$$Q_{stat} = \frac{\sum_{i=1}^n (Y_t - \bar{Y})(Y_{t-k} - \bar{Y})}{\sum (Y_t - \bar{Y})^2}$$

$$Q_{stat} = \frac{0,25233}{0,2994} = 0,8428$$

2. $Q_{stat} < X_n^2$ On accepte H_0 . Ce qui permet d'accepter l'hypothèse que la série est un processus Bruit Blanc. Donc, la série est stationnaire.

3. Le corrélogramme de la série (le graphe) montre que l'ensemble des valeurs (corrélogramme simple et corrélogramme partiel) sont à l'intérieur de l'intervalle de confiance. L'ensemble des valeurs sont non significatives. La série proposée est un bruit blanc. Donc la série est stationnaire.

Solution de l'exercice n°2

Le calcul de Q-stat pour h=4

$$Q - stat = n(n+2) \sum_{k=1}^4 \frac{r_k^2}{n-k}$$

$$Q - stat = 30 * 32 \left[\frac{(0,900)^2}{29} + \frac{(0,814)^2}{28} + \frac{(0,726)^2}{27} + \frac{(0,638)^2}{26} \right] = 83,26$$

1. 2. Le test de Box-pierces et de Lyung Box.

- $H_0 : \rho_1 = \rho_2 = \dots \rho_i = 0$
- $H_1 : \text{Il existe au moins un coefficient } \rho_i \neq 0$

$Q - stat > \chi_4^{0,95} = 9,48$ on rejette H_0 et on accepte H_1 . Il existe au moins un coefficient d'autocorrélation significativement différent de 0. Ce qui permet de dire que la série n'est pas stationnaire.

Solution de l'exercice n°3

1. Coefficient d'autocorrélation d'ordre k

$$r = \frac{cov(Y_t, Y_{t-k})}{\delta_{Y_t} \delta_{Y_{t-k}}} = \frac{\sum_{i=1}^n (Y_t - \bar{Y})(Y_{t-k} - \bar{Y}_{t-k})}{\sqrt{\sum (Y_t - \bar{Y})^2} * \sqrt{\sum (Y_{t-k} - \bar{Y}_{t-k})^2}}$$

Donc $r_1 = 0,021$, $r_2 = -0,502$, $r_3 = -0,348$

2. Calcul des intervalles de confiance et le test de significativité

En utilisant la formule de Barlett, on obtient :

$$V(r_1) = \frac{1}{n} = 0,5$$

$$V(r_2) = \frac{1}{n} (1 + 2r_1^2) = \frac{1}{10} (1 + 2(0,021)^2) = 0,1$$

$$V(r_3) = \frac{1}{n} (1 + 2(r_1^2 + r_2^2)) = \frac{1}{10} (1 + 2(0,021)^2 + (0,502)^2) = 0,15$$

Les intervalles de confiance :

$$k = 1: \left[-1,96 * \sqrt{\frac{1}{n}}; +1,96 * \sqrt{\frac{1}{n}} \right] = \left[-1,96 * \sqrt{0,1}; +1,96 * \sqrt{0,1} \right] = [-0,620; +0,620]$$

Le paramètre n'est pas significativement différent de 0 (la valeur du paramètre est à l'intérieur de l'intervalle de confiance).

$$k = 2: \left[-1,96 * \sqrt{0,1}; +1,96 * \sqrt{0,1} \right] = [-0,620; +0,620]$$

Le paramètre n'est pas significativement différent de 0 (la valeur du paramètre est à l'intérieur de l'intervalle de confiance).

$$k = 3: \left[-1,96 * \sqrt{0,15}; +1,96 * \sqrt{0,15} \right] = [-0,759; +0,759]$$

Le paramètre n'est pas significativement différent de 0 (la valeur du paramètre est à l'intérieur de l'intervalle de confiance).

3. Calcul des termes du corrélogramme partiel

En utilisant les estimateurs r_1 , r_2 et r_3 de ρ_1 , ρ_2 et ρ_3 On obtient :

$$r_{11} = r_1 = 0,021$$

$$\begin{bmatrix} r_{21} \\ r_{22} \end{bmatrix} = \begin{bmatrix} 1 & r_1 \\ r_1 & 1 \end{bmatrix}^{-1} \begin{bmatrix} r_1 \\ r_2 \end{bmatrix} = \begin{bmatrix} 1 & r_1 \\ r_1 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 0,021 \\ -0,502 \end{bmatrix} \rightarrow r_{22} = 0,502$$

$$\begin{bmatrix} r_{31} \\ r_{32} \\ r_{33} \end{bmatrix} = \begin{bmatrix} 1 & r_1 & r_2 \\ r_1 & 1 & r_1 \\ r_2 & r_1 & 1 \end{bmatrix}^{-1} \begin{bmatrix} r_1 \\ r_2 \\ r_3 \end{bmatrix} = \begin{bmatrix} 1 & 0,021 & -0,502 \\ 0,021 & 1 & 0,021 \\ -0,502 & 0,021 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 0,021 \\ -0,502 \\ -0,348 \end{bmatrix} = \begin{bmatrix} -0,185 \\ -0,489 \\ -0,431 \end{bmatrix} \text{ donc}$$

$$r_{33} = -0,431$$

4. Intervalle de confiance et test d'hypothèse

Sous l'hypothèse $H_0 : \rho_k = 0$ (pour $k=1, 2$ et 3) nous avons ρ_k suit une loi normale d'espérance nulle et de variance $\sqrt{\frac{1}{n}}$

Les intervalles de confiance au seuil de significativité de 95% sont égaux à

$$\left[-1,96 * \sqrt{\frac{1}{n}}; +1,96 * \sqrt{\frac{1}{n}} \right] = [-0,620; +0,620]$$

Aucun coefficient n'est à l'extérieur de cet intervalle, ils sont donc non significativement différent de 0.

5. Test de Ljung-Box pour $h=3$

- $H_0 : \rho_1 = \rho_2 = \dots \rho_i = 0$
- $H_1 : \text{Il existe au moins un coefficient } \rho_i \neq 0$

$$Q - stat = n(n+2) \sum_{k=1}^4 \frac{r_k^2}{n-k}$$

$$Q - stat = 10 * 12 \left[\frac{(0,021)^2}{9} + \frac{(-0,502)^2}{8} + \frac{(-0,348)^2}{7} \right] = 5,86$$

$Q_{stat} < X_n^2 = 7,8$ On accepte H_0 . Ce qui permet d'accepter l'hypothèse que la série est un processus Bruit Blanc. Donc, la série est stationnaire.