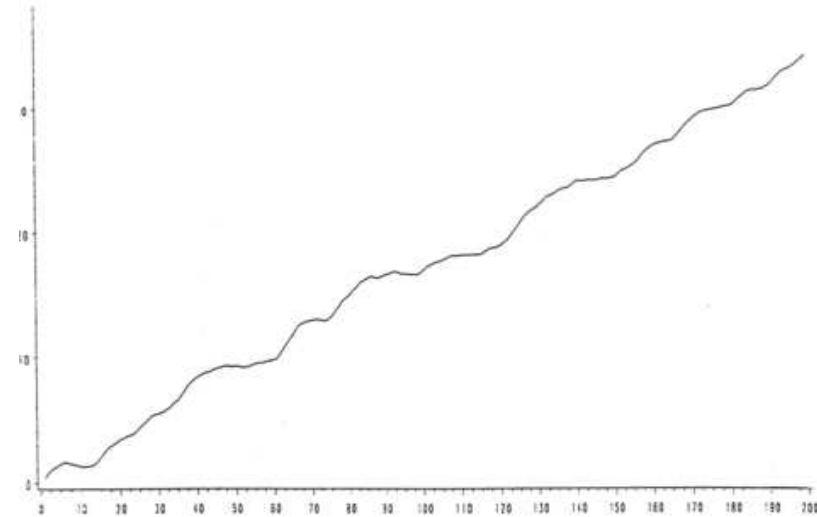


Séries temporelles linéaires
*Réponses question par question des travaux
dirigés n° 7*

Guillaume Lacôte
Bureau **E03**
✉ Guillaume.Lacote@ensae.fr
☞ <http://ensae.no-ip.com/SE206/>

Exercice corrigé 1

On considère une série réelle observée $(y_t)_{t \in [0, 200]}$ pour laquelle on cherche une représentation stationnaire, et dont la représentation est la suivante :



Justifier le modèle où y est la réalisation d'un processus Y non-stationnaire vérifiant

☞ Q1

$$\forall t \in \mathbb{Z}, \Phi(L)Y_t = \mu + \beta t + \epsilon_t$$

où ϵ est un bruit blanc de variance σ_ϵ^2 à déterminer.

L'observation de la série fait apparaître une tendance déterministe linéaire : la série observée est "assez proche" d'une certaine droite $\bar{y}_t = a + bt$.

Si l'on admet, comme le suggère son graphe, qu'à cette composante déterministe près elle est en outre stationnaire du second ordre, alors le processus Y dont elle est la réalisation admet une représentation auto-régressive infinie de la forme

$$\Phi(L)(Y_t - a - bt) = \epsilon_t$$

Or, à supposer que $\Phi(\mathbb{X}) = \sum_{k=0}^{+\infty} \phi_k \mathbb{X}^k$ ainsi que sa dérivée sont de rayon supérieur à 1 on a

$$\begin{aligned} \Phi(L) \circ (bt)_t &= b \sum_{k=0}^{+\infty} \phi_k (t-k) \\ &= b \left(\sum_{k=0}^{+\infty} \phi_k \right) t - b \left(\sum_{k=0}^{+\infty} k \phi_k \right) \\ &= b\Phi(1)t - b\Phi'(1) \end{aligned}$$

Comme $\Phi(L) \circ (a) = \Phi(1) \times a$ il vient en notant $\mu = a - b\Phi'(1)$ et $\beta = b\Phi(1)$

$$\Phi(L)Y_t = \mu + \beta t + \epsilon_t$$

En pratique, on se contente de choisir Φ de degré fini mais assez grand pour que ϵ soit significativement un bruit blanc.

Montrer qu'il existe un polynôme Φ^* tel que Y vérifie

$$\forall t \in \mathbb{Z}, \Delta Y_t = \mu + \beta t + \phi Y_{t-1} + \Phi^*(L)\Delta Y_t + \epsilon_t$$

où $\Delta = \mathbb{1} - L$.

Afin de tester la présence éventuelle d'une racine unitaire, on cherche à factoriser "de force" $\Phi(L)$ par $\mathbb{1} - L$ (transformation de BEVERIDGE-NELSON, voir TD 4, exercice 1) : pour ce faire on effectue la division euclidienne $\Phi(\mathbb{X})$ par $1 - \mathbb{X}$ ce qui s'écrit, comme 1 est racine de $\Phi(\mathbb{X}) - \Phi(1)$

$$\Phi^+(\mathbb{X}) = \frac{\Phi(\mathbb{X}) - \Phi(1)}{(1 - \mathbb{X})} \in \mathbb{R}[X], \quad d^\circ \Phi^+ \leq d^\circ \Phi - 1$$

Alors

$$\Phi(\mathbb{X}) = \Phi(1) + (1 - \mathbb{X})\Phi^+(\mathbb{X})$$

et donc

$$\begin{aligned} \Phi(L)Y &= \Phi(1) \times Y + (\mathbb{1} - L) \cdot \Phi^+(L) \circ Y \\ &= \Phi(1) \times Y + \Phi^+(L) \circ \Delta Y \\ &= \Phi(1) \times Y + \underbrace{\Phi^+(0)}_{\Phi(0) - \Phi(1)} \times \Delta Y + \underbrace{(\Phi^+(L) - \Phi^+(0))}_{-\Phi^*(L)} \circ \Delta Y \\ &= \Phi(1) \times Y + (1 - \Phi(1)) \times \Delta Y - \Phi^*(L) \circ \Delta Y \\ &= \Delta Y + \Phi(1)LY - \Phi^*(L) \circ \Delta Y \end{aligned}$$

et comme par ailleurs $\Phi(L)Y_t = \mu + \beta t + \epsilon_t$ il vient en définitive

¹Cette opération n'est normalement définie que pour les polynômes, i.e. lorsque Φ est de degré fini. Si $d^\circ \Phi = +\infty$ on définit $\Phi_N(\mathbb{X}) = \sum_{k=0}^N \phi_k \mathbb{X}^k$ dont on tire $\Phi_N^+(\mathbb{X})$ pour $N \geq 1$, dont on montre qu'elle est convergente de rayon supérieur à 1.

$$\Delta Y_t = \mu + \beta t + \Phi^*(L)\Delta Y_t - \Phi(1)Y_{t-1} + \epsilon_t$$

qui est bien le modèle recherché en posant $\phi = -\Phi(1)$.

Notons que par construction $\Phi^*(0) = 0$, donc $\Phi^*(\mathbb{X})$ ne contient que des puissances strictement positives de \mathbb{X} , et donc $\Phi^*(L) \circ \Delta Y_t$ ne contient que des valeurs passées $\Delta Y_{t-1}, \Delta Y_{t-2}, \dots, \Delta Y_t$.

Tester si le modèle est intégré, c'est tester si $\Phi(1) = 0$; sous cette forme cela revient juste à tester si le coefficient ϕ de Y_{t-1} est non-significativement non-nul.

Il aurait été possible de pratiquer une telle régression directement sur l'équation en Y_t au lieu de celle en ΔY_t ; mais comme Y n'est pas stationnaire, $Cov(Y_t, Y_{t+h})$ est significative pour tout h , donc la régression aurait dû comporter une infinité de variables explicatives $(Y_{t-h})_{h \geq 1}$ (l'inverse si ΔY est stationnaire, un nombre fini d'explicatives peut suffire).

On réalise alors la régression de ΔY_t sur $\langle 1, t, Y_{t-1}, \Delta Y_{t-1}, \dots, \Delta Y_{t-8} \rangle$; l'estimation des paramètres est la suivante (où LX désigne la série $(Y_{t-1})_{t \in \mathbb{Z}}$ et LY i désigne la série $(\Delta Y_{t-i})_{t \in \mathbb{Z}}$):

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob > F
Model	10	2.16857	0.21686	20.901	0.0001
Error	179	1.85724	0.01038		
C Total	189	4.02581			
Root MSE		0.10186	R-square	0.5387	
Dep Mean		0.17388	Adj R-sq	0.5129	
C.V.		58.58278			

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > T
INTERCEP	1	0.063825	0.02139593	2.983	0.0033
T	1	0.006102	0.00294649	2.982	0.0033
LX	1	-0.036870	0.01233214	-2.990	0.0032
LY1	1	0.727153	0.07300056	9.961	0.0001
LY2	1	-0.050103	0.09109175	-0.550	0.5830
LY3	1	-0.149478	0.09065962	-1.649	0.1009
LY4	1	-0.101392	0.09078450	-1.117	0.2656
LY5	1	-0.000012056	0.09050397	-0.000	0.9999
LY6	1	-0.043771	0.08985815	-0.487	0.6268
LY7	1	-0.000901	0.08964794	-0.010	0.9920
LY8	1	0.077040	0.07315238	1.053	0.2937

Q3

La modélisation proposée est-elle raisonnable?

Il est tout d'abord nécessaire de déterminer Φ tel que le modèle soit vérifié. Pour ce faire on choisit un degré arbitrairement grand pour Φ (ici 9)² et on vérifie que le modèle estimé par la régression des moindres carrés ordinaires³ est statistiquement valide; dans le cas présent R^2 ajusté (qui vaut 0,5129) assure que le modèle capture une part suffisante de la variance, donc que suffisamment de variables explicatives $\Delta Y_{t-1}, \Delta Y_{t-2}, \dots$ ont été introduites.

²Car $8 = d^\circ \Phi^* = d^\circ \Phi^+ = d^\circ \Phi - 1$.

³Puisque ϵ est censé être un bruit blanc.

Remarquons que lorsque ΔY est stationnaire, les coefficients de la régression sont donc asymptotiquement normaux, ce qui justifie l'emploi du test de Student pour tester leur significativité.⁴ En l'occurrence, on observe que les coefficients de la régression sur $\Delta Y_{t-4}, \dots, \Delta Y_{t-8}$ ne sont (individuellement) pas significativement non-nuls (colonne Prob > |t|).

Ainsi la modélisation proposée est valide mais le modèle estimé ici n'est pas significatif.

Q4 Justifier le modèle où y est la réalisation d'un processus Y stationnaire vérifiant

$$\forall t \in \mathbb{Z}, \Delta Y_t = \mu + \beta t + \phi Y_{t-1} + \phi_1 \Delta Y_{t-1} + \epsilon_t$$

On s'appuiera pour ce faire sur

Le test de $\phi_{\Delta Y_{t-4}} = \dots = \phi_{\Delta Y_{t-8}} = 0$ dans le modèle à 8 retards :

(a)

```

The SAS System
Dependent Variable: Y
Test:
      Numerator:    3.3081 DF:    5 F value:    0.7041
      Denominator: 0.010276 DF: 179 Prob>F:    0.5624
    
```

La régression de ΔY_t sur $\langle 1, t, Y_{t-1}, \Delta Y_{t-1}, \dots, \Delta Y_{t-3} \rangle$:

(b)

```

Analysis of Variance
Source      DF      Sum of Squares      Mean Square      F Value      Prob>F
Model            6      2.24814      0.37802      36.277      0.0001
Error           193      24.951      0.12927
Total          199      27.199
Root MSE      0.11373      R-square      0.3777
Dep Mean      0.16794      Adj R-sq      0.3531
C.V.          60.06933
Parameter Estimates
Variable    DF      Parameter Estimate      Standard Error      T for H0: Parameter=0      Prob > |t|
INTERCEP    1      0.062633      0.01847919      3.389      0.0009
t            1      0.005271      0.01877704      0.2808      0.0055
LY          1      -0.031451      0.01128994      -2.786      0.0059
LV1         1      0.740109      0.07085748      10.445      0.0001
LV2         1      -0.040810      0.08846408      -0.466      0.4938
LV3         1      0.144091      0.08863452      1.626      0.1057
    
```

Le test de $\phi_{\Delta Y_{t-2}} = \dots = \phi_{\Delta Y_{t-8}} = 0$ dans le modèle à 8 retards :

(c)

```

The SAS System
Dependent Variable: Y
Test:
      Numerator:    0.0068 DF:    7 F value:    0.6597
      Denominator: 0.010376 DF: 179 Prob>F:    0.7059
    
```

⁴En toute généralité il serait possible que ΔY ne soit pas stationnaire, auquel cas l'emploi du test de Student serait inadapté.

La régression de ΔY_t sur $\langle 1, t, Y_{t-1}, \Delta Y_{t-1} \rangle$:

(d)

```

Analysis of Variance
Source      DF      Sum of Squares      Mean Square      F Value      Prob>F
Model            3      2.22742      0.74247      71.308      0.0001
Error           193      2.00731      0.01040
Total          196      4.23473
Root MSE      0.10178      R-square      0.5260
Dep Mean      0.17070      Adj R-sq      0.5186
C.V.          59.74365
Parameter Estimates
Variable    DF      Parameter Estimate      Standard Error      T for H0: Parameter=0      Prob > |t|
INTERCEP    1      0.057473      0.01745812      3.292      0.0012
t            1      0.005824      0.00171041      3.405      0.0008
LY          1      -0.034680      0.01028056      -3.373      0.0009
LV1         1      0.719836      0.04962822      14.505      0.0001
    
```

Les auto-corrélations directes du modèle final

(e)

```

Autocorrelations
Lag Covariance Correlation
0 0.010189 1.00000
1 0.0014062 0.13800
2 -0.0004064 -0.03989
3 -0.0011716 -0.11498
4 -0.0020711 -0.20322
5 -0.0001957 -0.01901
6 -0.0004443 -0.04382
7 -0.0003036 -0.02988
8 0.00047225 0.04635
9 0.0003036 0.02988
10 -0.0008979 -0.00961
11 0.0008549 0.08391
12 0.0008212 0.08010
13 -0.0007829 -0.07684
14 0.00052963 0.05217
15 0.00058924 0.05785
16 -0.0004967 -0.04829
17 0.00019926 0.01956
18 -0.0013391 -0.13162
19 0.00028753 0.02822
20 0.0000309 0.00304
21 -0.0006792 -0.06666
22 0.0000193 0.00195
23 -0.0006871 -0.06566
24 0.0002936 0.02741
*** marks two standard errors
    
```

Les auto-corrélations inverses du modèle final

(f)

```

Inverse Autocorrelations
Lag Correlation
1 -0.05452
2 0.03347
3 0.13305
4 0.22247
5 0.30644
6 0.35505
7 0.35189
8 -0.03374
9 -0.06734
10 -0.02248
11 -0.04162
12 0.02220
13 0.07864
14 -0.00582
15 -0.00996
16 0.00537
17 -0.02800
18 0.12947
19 -0.00012
20 -0.01547
21 0.02619
22 -0.09869
23 0.08316
24 -0.04172
    
```

Les auto-corrélations partielles du modèle final

Lag	Correlation	-1	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	1	
1	0.01380																						
2	-0.04008											*											
3	-0.11632												**										
4	-0.02618												*										
5	-0.00875												*										
6	-0.05951												*										
7	-0.04355												*										
8	0.04756												*										
9	0.03607												*										
10	0.00167												*										
11	0.07362												*										
12	-0.00746												*										
13	-0.07370												*										
14	0.02050												*										
15	0.06359												*										
16	0.01179												*										
17	-0.02095												*										
18	-0.14741												**										
19	0.02405												*										
20	-0.01415												*										
21	-0.01533												*										
22	0.09851												**										
23	-0.08730												**										
24	0.04548												*										

(g)

Le test de Porte-Manteau du bruit blanc du modèle final :

Lag	Chi Square	DF	Prob	Autocorrelations																				
5	3.58	6	0.733	0.014	-0.040	0.115	-0.020	-0.019	-0.044															
12	6.22	12	0.905	-0.049	0.046	0.029	-0.010	0.084	0.006															
18	12.30	18	0.831	-0.077	0.032	0.058	-0.009	0.020	-0.131															
24	16.75	24	0.859	0.028	0.009	-0.067	0.097	-0.066	0.027															

(h)

On se propose donc de déterminer le sous-modèle minimal qui soit totalement significatif.
Remarque : Bien que l'on observe à ce stade que le coefficient en Y_{t-1} est significativement non-nul, **on ne peut pas** en déduire qu'il le sera aussi dans le sous-modèle où tous les coefficients sont significatifs.⁵ Il est nécessaire de déterminer ce sous-modèle et d'y tester la significativité du coefficient en Y_{t-1} .

- Compte-tenu de la régression précédente, on teste la nullité **simultanée** des coefficients de $\Delta Y_{t-4}, \dots, \Delta Y_{t-8}$ dans la régression. En l'occurrence, l'hypothèse $\mathcal{H}_0 : \phi_{\Delta Y_{t-4}} = \dots = \phi_{\Delta Y_{t-8}} = 0$ est acceptée car la statistique de Fisher associée vaut **0,7841**.

⁵La statistique de Student du coefficient β dans la régression ordinaire

$$Y = \alpha X^1 + \beta X^2 + \gamma X^3 + U$$

est

$$t_{X^2} = \frac{\hat{\beta} - \beta_0}{\sqrt{\hat{\sigma}_2^2}}$$

où $\hat{\sigma}_2^2$ désigne le terme diagonal de la matrice de variance-covariance Σ de $\begin{pmatrix} \alpha \\ \beta \\ \gamma \end{pmatrix}$.

Or rien ne dit que cette matrice ne dépend pas de X^3 , même si γ est non-significativement non-nul : comme $\hat{\beta} = (X^2' X^2)^{-1} X^2' Y$ et $\hat{\gamma} = (X^3' X^3)^{-1} X^3' Y$ pour peu que $Cov(X^2, X^3) \neq 0$, la matrice Σ dépendra de $\hat{\gamma}$ (au sens où $Cov(\Sigma^{1,3}, \hat{\gamma}) \neq 0$).

En l'occurrence, on sait (voir TD 2, exercice 3) que ΔY est stationnaire que ΔY_{t-1} et ΔY_{t-i} sont certainement corrélés.

- On effectue donc la régression de ΔY sur $\langle 1, t, Y_{t-1}, \Delta Y_{t-1}, \dots, \Delta Y_{t-4} \rangle$, et on observe que le modèle ainsi estimé est valide (ce qui est heureux). Il apparaît cependant que les coefficients de ΔY_{t-2} et de ΔY_{t-3} sont **individuellement** non-significativement non-nuls.
- On effectue donc le test de la nullité simultanée de ces coefficients en formulant l'hypothèse $\mathcal{H}_0 : \phi_{\Delta Y_{t-2}} = \dots = \phi_{\Delta Y_{t-8}} = 0$.

Remarque : Noter que tester successivement

- la nullité simultanée des coefficients de $\Delta Y_{t-4}, \dots, \Delta Y_{t-8}$ dans la régression sur $\langle 1, t, Y_{t-1}, \Delta Y_{t-1}, \dots, \Delta Y_{t-8} \rangle$
- puis la nullité simultanée des coefficients de ΔY_{t-2} et ΔY_{t-3} dans la régression sur $\langle 1, t, Y_{t-1}, \Delta Y_{t-1}, \dots, \Delta Y_{t-3} \rangle$

est moins robuste statistiquement que de tester directement la nullité simultanée des coefficients de $\Delta Y_{t-2}, \dots, \Delta Y_{t-8}$ dans la régression sur $\langle 1, t, Y_{t-1}, \Delta Y_{t-1}, \dots, \Delta Y_{t-8} \rangle$.

En l'occurrence, l'hypothèse est acceptée car la statistique de Fisher associée vaut **0,6597**

- On effectue donc en définitive la régression de ΔY sur $\langle 1, t, Y_{t-1}, \Delta Y_{t-1} \rangle$, et on observe que le modèle ainsi estimé est valide. On constate en outre qu'aucun coefficient n'est individuellement non-significatif.

On procède donc à l'identification du modèle AR (voir TD 5, exercice 1) en ΔY sur les explicatives t et LY .

On observe en l'occurrence que les auto-corrélations directes décroissent "exponentiellement" vers 0 (en l'occurrence, elles ne sont plus significativement non-nulles dès l'ordre $h \geq 3$); en outre le test de Porte-Manteau permet de conclure que le résidu estimé est non-significativement différent d'un bruit blanc.

Q5

Montrer que le modèle retenu s'écrit

$$(\mathbb{1} - \rho L)Y_t = \mu + \beta t + \phi_1 \Delta Y_{t-1} + \epsilon_t$$

et exprimer ρ en fonction de ϕ .

Le modèle estimé s'écrit

$$\Delta Y_t = \mu + \beta t + \phi Y_{t-1} + \phi_1 \Delta Y_{t-1} + \epsilon_t$$

où ϕ_1 est le coefficient en X de $\Phi^*(X)$, ce qui se réécrit encore

$$(\mathbb{1} - \rho L)Y_t = \mu + \beta t + \phi_1 \Delta Y_{t-1} + \epsilon_t$$

en posant $\rho = 1 + \phi$.

Q6 Tester l'hypothèse " $\mathcal{H}_0^1 : \beta = 0$ et $\rho = 1$ " en utilisant les tables suivantes :

(a)	Dependent Variable: Y	F value: 5.8394
Test:	Numerator: 0.0607 DF: 2	Prob>F: 0.0039
	Denominator: 0.010401 DF: 193	

(b)

Sample size n	Probability of a smaller value			
	0.90	0.95	0.975	0.99
25	2.20	3.61	2.97	3.41
50	2.18	2.56	2.39	3.28
100	2.17	2.54	2.36	3.22
250	2.16	2.53	2.34	3.19
500	2.16	2.52	2.33	3.18
∞	2.16	2.52	2.33	3.18
s.c.	0.003	0.004	0.006	0.008

Sample size n	Probability of a smaller value			
	0.90	0.95	0.975	0.99
25	2.77	3.20	3.59	4.05
50	2.75	3.14	3.47	3.87
100	2.73	3.11	3.42	3.78
250	2.73	3.09	3.39	3.74
500	2.73	3.08	3.38	3.72
∞	2.72	3.08	3.38	3.71
s.c.	0.004	0.005	0.007	0.008

(c)

Sample size n	Probability of a smaller value			
	0.90	0.95	0.975	0.99
25	2.39	2.85	3.25	3.74
50	2.38	2.81	3.18	3.60
100	2.38	2.79	3.14	3.53
250	2.38	2.79	3.12	3.49
500	2.38	2.78	3.11	3.48
∞	2.38	2.78	3.11	3.48
s.c.	0.004	0.005	0.006	0.009

Sample size n	Probability of a smaller value							
	0.01	0.025	0.05	0.10	0.90	0.95	0.975	0.99
25	0.61	0.75	0.89	1.10	4.67	5.68	6.75	8.21
50	0.62	0.77	0.91	1.12	4.31	5.13	5.94	7.02
100	0.63	0.77	0.92	1.12	4.16	4.88	5.59	6.50
250	0.63	0.77	0.92	1.13	4.07	4.75	5.40	6.22
500	0.63	0.77	0.92	1.13	4.05	4.71	5.35	6.15
∞	0.63	0.77	0.92	1.13	4.03	4.68	5.31	6.09
s.c.	0.003	0.003	0.003	0.003	0.01	0.02	0.03	0.05

(d)

Sample size n	Probability of a smaller value							
	0.01	0.025	0.05	0.10	0.90	0.95	0.975	0.99
25	0.29	0.38	0.49	0.65	4.12	5.18	6.30	7.88
50	0.29	0.39	0.50	0.66	3.94	4.86	5.80	7.06
100	0.29	0.39	0.50	0.67	3.86	4.71	5.57	6.70
250	0.30	0.39	0.51	0.67	3.81	4.63	5.45	6.52
500	0.30	0.39	0.51	0.67	3.79	4.61	5.41	6.47
∞	0.30	0.40	0.51	0.67	3.78	4.59	5.38	6.43
s.c.	0.002	0.002	0.002	0.002	0.01	0.02	0.03	0.05

Sample size n	Probability of a smaller value							
	0.01	0.025	0.05	0.10	0.90	0.95	0.975	0.99
25	0.74	0.90	1.08	1.33	5.91	7.34	8.65	10.61
50	0.76	0.93	1.11	1.37	5.61	6.73	7.81	9.31
100	0.76	0.94	1.12	1.38	5.47	6.49	7.44	8.73
250	0.76	0.94	1.13	1.39	5.39	6.34	7.25	8.43
500	0.76	0.94	1.13	1.39	5.36	6.30	7.20	8.34
∞	0.77	0.94	1.13	1.39	5.34	6.25	7.16	8.27
s.c.	0.004	0.004	0.003	0.004	0.015	0.020	0.032	0.058

(e)

Sample size T	Probability that $T(\hat{\beta} - 1)$ is less than entry							
	0.01	0.025	0.05	0.10	0.90	0.95	0.975	0.99
Case 1								
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
∞	-13.8	-10.5	-8.1	-5.7	0.93	1.28	1.60	2.03
Case 2								
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
∞	-20.7	-16.9	-14.1	-11.3	-0.85	-0.13	0.41	1.04
Case 4								
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	-0.80
500	-28.9	-24.8	-21.5	-18.1	-3.76	-2.65	-1.78	-0.80
∞	-29.5	-25.1	-21.8	-18.3	-3.77	-2.66	-1.79	-0.81
The probability shown at the head of the column is the area in the left-hand tail.								
Source: Wayne A. Fuller, <i>Introduction to Statistical Time Series</i> , Wiley, New York, 1976, p. 371.								

Sample size T	Probability that $(\hat{\beta} - 1)/\hat{\sigma}_e$ is less than entry							
	0.01	0.025	0.05	0.10	0.90	0.95	0.975	0.99
Case 1								
25	-2.66	-2.26	-1.95	-1.60	0.92	1.33	1.70	2.08
50	-2.62	-2.25	-1.95	-1.61	0.91	1.31	1.66	2.04
100	-2.60	-2.24	-1.95	-1.61	0.90	1.29	1.64	2.02
250	-2.58	-2.23	-1.95	-1.62	0.89	1.29	1.63	2.01
500	-2.58	-2.23	-1.95	-1.62	0.89	1.28	1.62	2.01
∞	-2.58	-2.23	-1.95	-1.62	0.89	1.28	1.62	2.01
Case 2								
25	-3.75	-3.33	-3.00	-2.63	-0.37	0.00	0.34	0.94
50	-3.58	-3.22	-2.93	-2.60	-0.40	-0.03	0.29	0.87
100	-3.51	-3.17	-2.89	-2.58	-0.42	-0.05	0.26	0.82
250	-3.46	-3.14	-2.88	-2.57	-0.42	-0.06	0.24	0.80
500	-3.44	-3.13	-2.87	-2.57	-0.43	-0.07	0.24	0.80
∞	-3.43	-3.12	-2.86	-2.57	-0.44	-0.07	0.23	0.79
Case 4								
25	-4.38	-3.95	-3.60	-3.24	-1.14	-0.80	-0.50	0.00
50	-4.15	-3.80	-3.50	-3.18	-1.19	-0.87	-0.58	-0.29
100	-4.04	-3.73	-3.45	-3.15	-1.22	-0.90	-0.62	-0.34
250	-3.99	-3.69	-3.43	-3.13	-1.23	-0.92	-0.64	-0.36
500	-3.98	-3.68	-3.42	-3.13	-1.24	-0.93	-0.66	-0.37
∞	-3.96	-3.66	-3.41	-3.12	-1.25	-0.94	-0.66	-0.37
The probability shown at the head of the column is the area in the left-hand tail.								
Source: Wayne A. Fuller, <i>Introduction to Statistical Time Series</i> , Wiley, New York, 1976.								

Sous \mathcal{H}_0^1 le modèle se réécrit

$$Y_t = \mu + Y_{t-1} + \phi_1 \Delta Y_{t-1} + \epsilon_t$$

La statistique de test associée est celle de la table VI, selon laquelle le fractile à 5% vaut 5,83 (on retient la valeur pour 250 observations⁶); comparée à la valeur empirique F-Value : 5,83 on accepte donc l'hypothèse \mathcal{H}_0^1 .

⇨ Q7 Tester finalement l'hypothèse " $\mathcal{H}_0^2 : \mu = 0, \beta = 0$ et $\rho = 1$ " et conclure.

On constate que $\mathcal{H}_0^2 \rightarrow \mathcal{H}_0^1$; comme \mathcal{H}_0^1 est acceptable on se propose de tester \mathcal{H}_0^2 .

Sous \mathcal{H}_0^2 le modèle se réécrit

$$Y_t = Y_{t-1} + \phi_1 \Delta Y_{t-1} + \epsilon_t$$

La statistique de test associée est celle de la table V, selon laquelle le fractile à 5% vaut 4,75 comparée à la valeur empirique F-Value : 10,75 on rejette donc l'hypothèse \mathcal{H}_0^2 .

Exercice corrigé 2

On considère un processus stationnaire multivarié $X \in ((\mathbb{R}^n)^\Omega)^Z$, $n \geq 2$, qui suit le modèle auto-régressif

$$\Phi(L)X = \epsilon$$

⁶Plutôt que celle pour 100 observations, car le test est ainsi plus restrictif : si on accepte \mathcal{H}_0 avec le fractile à 100 observations, on l'accepterait a fortiori avec celui à 100 observations.

où ϵ est un bruit blanc de variance-covariance Σ supposée définie positive. On suppose en outre que le modèle est sous forme canonique, c'est-à-dire que les racines de $Det\Phi$ sont toutes de module strictement supérieur à un.

On se donne $m \in \llbracket 1, n-1 \rrbracket$ et on note $X = \begin{pmatrix} X^1 \\ X^2 \end{pmatrix} \begin{matrix} \uparrow m \\ \downarrow n-m \end{matrix}$; on cherche à donner un sens à l'expression "y est la cause x".⁷

Dans la suite pour toute matrice A de $\mathcal{M}_n(\mathbb{R})$ sera notée $A = \left(\begin{array}{c|c} A^{1,1} & A^{1,2} \\ \hline A^{2,1} & A^{2,2} \end{array} \right) \begin{matrix} \uparrow m \\ \downarrow n-m \end{matrix}$.

Partie 1 Définition de la notion de causalité

Q1 A un instant donné, l'intuition suggère que si X_t^2 cause X_t^1 alors X_t^2 intervient significativement dans la valeur de X_t^1 ; en particulier la prévision optimale de X_t^1 n'est dans ce cas pas la même selon que l'on connaît X_t^2 ou pas. On dit donc que X_t^2 ne cause pas instantanément X_t^1 au sens de Granger, et on note $X^2 \not\rightsquigarrow X^1$, ssi $\mathbb{E}\mathbb{L}(X_t^1|X_t^2, X_{t-1}, \dots) = \mathbb{E}\mathbb{L}(X_t^1|X_{t-1}, \dots)$

(a) $\mathbb{E}\mathbb{L}(X_t^1|X_t^2, X_{t-1}, \dots) = \mathbb{E}\mathbb{L}(X_t^1|X_{t-1}, \dots) + \mathbb{E}\mathbb{L}(\epsilon_t^1|\epsilon_t^2)$

Le modèle étant sous forme canonique, ϵ est l'innovation du processus vectoriel X : par définition de l'innovation $\epsilon_t = X_t - \mathbb{E}\mathbb{L}(X_t|X_{t-1}, \dots)$ et donc en projetant pour $i \in \{1, 2\}$

$$\epsilon_t^i = X_t^i - \mathbb{E}\mathbb{L}(X_t^i|X_{t-1}, \dots)$$

(ϵ^2 est l'innovation du processus $X^2|X^1$)

$$\begin{aligned} \mathbb{E}\mathbb{L}(X_t^1|X_t^2, X_{t-1}, \dots) &= \mathbb{E}\mathbb{L}(X_t^1|(X_t^2, X_{t-1}^2, \dots), (X_{t-1}^1, \dots)) \\ &= \mathbb{E}\mathbb{L}(X_t^1|(\epsilon_t^2, \epsilon_{t-1}^2, \dots), (X_{t-1}^1, \dots)) \\ &= \mathbb{E}\mathbb{L}(X_t^1|\epsilon_t^2, (X_{t-1}^2, \dots, X_{t-1}^1, \dots)) \\ &= \mathbb{E}\mathbb{L}(X_t^1|\epsilon_t^2, X_{t-1}, \dots) \end{aligned}$$

Or $\forall t \in \mathbb{Z}$, $\epsilon_t^2 \perp\!\!\!\perp \langle X_{t-1}, \dots \rangle$ donc $\langle \epsilon_t^2, X_{t-1}, \dots \rangle = \langle \epsilon_t^2 \rangle \oplus \langle X_{t-1}, \dots \rangle$ et donc

$$\mathbb{E}\mathbb{L}(X_t^1|X_t^2, X_{t-1}, \dots) = \mathbb{E}\mathbb{L}(X_t^1|\epsilon_t^2) \oplus \mathbb{E}\mathbb{L}(X_t^1|X_{t-1}, \dots)$$

Or

$$X_t = \sum_{k=1}^{d^\circ\Phi} \Phi_k X_{t-k} + \epsilon_t$$

et $\epsilon_t \perp\!\!\!\perp \langle X_{t-1}, \dots \rangle$ donc en projetant

$$\mathbb{E}\mathbb{L}(X_t^1|\epsilon_t^2) = \mathbb{E}\mathbb{L}(\epsilon_t^1|\epsilon_t^2)$$

et donc finalement

⁷La démarche proposée est tirée (de façon très simplifiée) des travaux de Clive W.J. Granger, co-lauréat avec Robert F. Engle du prix Nobel d'économie 2003. C. Granger fut lauréat non pas pour la notion de causalité qu'il définit (utilisée pour établir la thèse polémique du réchauffement de la planète), mais pour ses résultats de cointégration. Voir <http://www.nobel.se/economics/laureates/2003/ecoadv.pdf>

$$\mathbb{E}\mathbb{L}(X_t^1|X_t^2, X_{t-1}, \dots) = \mathbb{E}\mathbb{L}(X_t^1|X_{t-1}, \dots) + \mathbb{E}\mathbb{L}(\epsilon_t^1|\epsilon_t^2)$$

Montrer que

$$X_t^2 \not\rightsquigarrow X_t^1 \Leftrightarrow \Sigma^{1,2} = (0) \Leftrightarrow X_t^1 \not\rightsquigarrow X_t^2$$

(b) (on dit dans ce cas par symétrie que X^2 et X^1 ne se causent pas instantanément au sens de Granger).

Par définition pour $t \in \mathbb{Z}$

$$\begin{aligned} X_t^2 \not\rightsquigarrow X_t^1 &\Leftrightarrow \mathbb{E}\mathbb{L}(X_t^1|X_t^2, X_{t-1}, \dots) = \mathbb{E}\mathbb{L}(X_t^1|X_{t-1}, \dots) \\ &\Leftrightarrow \mathbb{E}\mathbb{L}(\epsilon_t^1|\epsilon_t^2) = 0 \\ &\Leftrightarrow \text{Cov}(\epsilon_t^1, \epsilon_t^2) = (0) \\ &\Leftrightarrow \Sigma^{1,2} = (0) \\ &\Leftrightarrow \text{Cov}(\epsilon_t^1, \epsilon_t^2) = (0) \\ &\Leftrightarrow \mathbb{E}\mathbb{L}(\epsilon_t^2|\epsilon_t^1) = 0 \\ &\Leftrightarrow \mathbb{E}\mathbb{L}(X_t^2|X_t^1, X_{t-1}, \dots) = \mathbb{E}\mathbb{L}(X_t^2|X_{t-1}, \dots) \\ &\Leftrightarrow X_t^1 \not\rightsquigarrow X_t^2 \end{aligned}$$

Q2 Par extension on dit alors que X^2 ne cause pas globalement X^1 , ou simplement que ne cause pas X^1 au sens de Granger (ce que l'on note $X^2 \not\rightsquigarrow X^1$) ssi

$$\forall t \in \mathbb{Z}, \mathbb{E}\mathbb{L}(X_t^1|X_{t-1}, \dots) = \mathbb{E}\mathbb{L}(X_t^1|X_{t-1}^1, \dots)$$

On cherche à montrer que

$$X^2 \not\rightsquigarrow X^1 \Leftrightarrow \Phi(\mathbb{X})^{1,2} = (0)$$

(où $\Phi(\mathbb{X})^{1,2}$ désigne le bloc supérieur droit du polynôme matriciel $\Phi(\mathbb{X})$).

(a) $\text{Montrer tout d'abord que } \Phi(\mathbb{X})^{1,2} = (0) \rightarrow X^2 \not\rightsquigarrow X^1$.

On a pour $t \in \mathbb{Z}$

$$X_t = \sum_{k=1}^{d^\circ\Phi} \phi_k X_{t-k} + \epsilon_t$$

et donc en projetant

$$X_t^1 = \sum_{k=1}^{d^\circ\Phi} \phi_k^{1,1} X_{t-k}^1 + \sum_{k=1}^{d^\circ\Phi} \phi_k^{1,2} X_{t-k}^2 + \epsilon_t^1$$

Si $\Phi(\mathbb{X})^{1,2} = (0)$ alors

$$X_t^1 = \sum_{k=1}^{d^\circ\Phi} \phi_k^{1,1} X_{t-k}^1 + \epsilon_t^1$$

et donc

$$\mathbb{E}\mathbb{L}(X_t^1 | X_{t-1}, \dots) = \sum_{k=1}^{d^\circ \Phi} \phi_k^{1,1} X_{t-k}^1 = \boxed{\mathbb{E}\mathbb{L}(X_t^1 | X_{t-1}^1, \dots)}$$

c'est-à-dire $X^2 \not\approx X^1$.

(b) Ecrire dans le cas général la décomposition de Wold de X^1 sur $\langle \epsilon^1, \epsilon^2 \rangle$.

Comme ϵ est l'innovation de X , soit $(A_k)_{k \in \mathbb{N}} \in \mathcal{M}_n(\mathbb{R})^{\mathbb{N}}$ telle que

$$\forall t \in \mathbb{Z}, X_t = \sum_{k=0}^{+\infty} A_k \epsilon_{t-k}$$

Alors en projetant il vient ⁸

$$\forall t \in \mathbb{Z}, X_t^1 = \sum_{k=0}^{+\infty} A_k^{1,1} \epsilon_{t-k}^1 + \sum_{k=0}^{+\infty} A_k^{1,2} \epsilon_{t-k}^2$$

(c) On suppose que $X^2 \not\approx X^1$; montrer qu' ϵ^1 est alors l'innovation de X^1 .

Supposons alors que $X^2 \approx X^1$; alors pour tout $t \in \mathbb{Z}$ $\mathbb{E}\mathbb{L}(X_t^1 | X_{t-1}, \dots) = \mathbb{E}\mathbb{L}(X_t^1 | X_{t-1}^1, \dots)$ et donc $X_t^1 - \mathbb{E}\mathbb{L}(X_t^1 | X_{t-1}^1, \dots) = X_t^1 - \mathbb{E}\mathbb{L}(X_t^1 | X_{t-1}, \dots) = \epsilon_t^1$; ϵ^1 est donc l'innovation de X^1 .

(d) En déduire que si $X^2 \not\approx X^1$, alors $\Phi(\mathbb{X})^{1,2} = (0)$.

Si $X^2 \not\approx X^1$, alors X^1 admet une représentation de Wold sous la forme

$$X_t^1 = \sum_{k=0}^{+\infty} B_k \epsilon_{t-k}^1$$

où $B_k \in \mathcal{M}_m(\mathbb{R})$. On a par ailleurs pour tout $t \in \mathbb{Z}$

$$X_t^2 = \sum_{k=0}^{+\infty} A_k^{2,1} \epsilon_{t-k}^1 + \sum_{k=0}^{+\infty} A_k^{2,2} \epsilon_{t-k}^2$$

et donc

$$\begin{pmatrix} X_t^1 \\ X_t^2 \end{pmatrix} = \begin{pmatrix} \sum_{k=0}^{+\infty} B_k & (0) \\ \sum_{k=0}^{+\infty} A_k^{2,1} & \sum_{k=0}^{+\infty} A_k^{2,2} \end{pmatrix} \times \begin{pmatrix} \epsilon_{t-k}^1 \\ \epsilon_{t-k}^2 \end{pmatrix}$$

⁸Chacune des deux sommes converge car $\|A_k^{1,j}\|_2^2 \leq \|A_k\|_2^2$ et que $\sum_k A_k$ est normalement convergente.

soit ⁹

$$X = \sum_{k=0}^{+\infty} \begin{pmatrix} B_k & (0) \\ A_k^{2,1} & A_k^{2,2} \end{pmatrix} L^k \circ \epsilon$$

Alors

$$\epsilon = \Phi(L)X = \Phi(L) \left(\sum_{k=0}^{+\infty} \begin{pmatrix} B_k & (0) \\ A_k^{2,1} & A_k^{2,2} \end{pmatrix} L^k \right) \circ \epsilon$$

et donc la matrice $\left(\sum_{k=0}^{+\infty} \begin{pmatrix} B_k & (0) \\ A_k^{2,1} & A_k^{2,2} \end{pmatrix} \mathbb{X}^k \right)$ sur l'anneau $\mathcal{M}_n(\mathbb{R})[\mathbb{X}]$ est régulière et est une inverse de $\Phi(\mathbb{X})$, ce qui s'écrit encore par blocs

$$\begin{pmatrix} \Phi(\mathbb{X})^{1,1} & \Phi(\mathbb{X})^{1,2} \\ \Phi(\mathbb{X})^{2,1} & \Phi(\mathbb{X})^{2,2} \end{pmatrix} \times \begin{pmatrix} \left(\sum_{k=0}^{+\infty} B_k \mathbb{X}^k \right) & (0) \\ \left(\sum_{k=0}^{+\infty} A_k^{2,1} \mathbb{X}^k \right) & \left(\sum_{k=0}^{+\infty} A_k^{2,2} \mathbb{X}^k \right) \end{pmatrix} = \begin{pmatrix} Id_m & (0) \\ (0) & Id_{n-m} \end{pmatrix}$$

et en particulier en considérant le bloc (1, 2) il vient

$$(0) \times \Phi(\mathbb{X})^{1,1} + \left(\sum_{k=0}^{+\infty} A_k^{2,2} \mathbb{X}^k \right) \times \Phi(\mathbb{X})^{1,2} = (0)$$

Or $\begin{pmatrix} \sum_{k=0}^{+\infty} B_k \mathbb{X}^k & (0) \\ \sum_{k=0}^{+\infty} A_k^{2,1} \mathbb{X}^k & \sum_{k=0}^{+\infty} A_k^{2,2} \mathbb{X}^k \end{pmatrix}$ est inversible, donc $\left(\sum_{k=0}^{+\infty} A_k^{2,2} \mathbb{X}^k \right)$ également ¹⁰ donc

$\Phi(\mathbb{X})^{1,2} = (0) \in \mathcal{M}_n(\mathbb{R})[\mathbb{X}]$ et finalement

$$\boxed{\Phi(L)^{1,2} = (0)}$$

ce qui achève la preuve de la réciproque.

■ Partie 2 Test de la notion de causalité

On se place dans le cas où $m = n - m = 1$, et on note $p = d^\circ \Phi$.

On suppose de plus que $\Phi(0) = Id$.

⁹Avec la convention que

$$ML^k = \begin{pmatrix} M^{1,1}L^k & \dots & M^{1,n}L^k \\ \vdots & & \vdots \\ M^{n,1}L^k & \dots & M^{n,n}L^k \end{pmatrix}$$

et en remarquant que chaque $\sum_k A_k^{i,j} \mathbb{X}^k$ est convergente car $|A_k^{i,j}| \leq \|A_k\|_2$ et que $\sum_k A_k$ est normalement convergente

¹⁰Comme la matrice $\begin{pmatrix} A & 0 \\ B & C \end{pmatrix}$ est triangulaire par blocs, si elle est inversible, alors C l'est également.

Montrer que X vérifie aussi le modèle

$$\Psi(L)X = \eta$$

pour un certain polynôme matriciel Ψ (de degré p) et un bruit blanc η à déterminer

On sait que

$$\Phi(L)X = \epsilon$$

donc pour tout $\alpha \in \mathbb{R}$ en multipliant par $\begin{pmatrix} 1 & 0 \\ -\alpha & 1 \end{pmatrix}$ il vient

$$\left(\begin{pmatrix} 1 & 0 \\ -\alpha & 1 \end{pmatrix} \times \Phi(L) \right) \circ X = \begin{pmatrix} 1 & 0 \\ -\alpha & 1 \end{pmatrix} \times \epsilon$$

Notons donc $\eta = \begin{pmatrix} 1 & 0 \\ -\alpha & 1 \end{pmatrix} \times \epsilon$; alors pour tout $t \in \mathbb{Z}$ et $l \in \mathbb{Z}$

$$\text{Cov}(\eta_t, \eta_{t-l}) = \begin{pmatrix} 1 & 0 \\ -\alpha & 1 \end{pmatrix}' \text{Cov}(\epsilon_t, \epsilon_{t-l}) \begin{pmatrix} 1 & 0 \\ -\alpha & 1 \end{pmatrix}$$

Posons en particulier $\alpha = \frac{\Sigma^{1,2}}{\Sigma^{1,1}}$: alors $\text{Cov}(\eta_t, \eta_{t-l}) = 0$ et donc η est un bruit blanc. Sa matrice de variance-covariance est alors

$$\begin{pmatrix} 1 & 0 \\ -\alpha & 1 \end{pmatrix} \Sigma \begin{pmatrix} 1 & -\alpha \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} \Sigma^{1,1} & 0 \\ 0 & \Sigma^{2,2} - \frac{(\Sigma^{1,2})^2}{\Sigma^{1,1}} \end{pmatrix}$$

Notons enfin $\Psi(\mathbb{X})$ la matrice à coefficients dans $\mathbb{R}[\mathbb{X}]$ $\begin{pmatrix} 1 & 0 \\ -\alpha & 1 \end{pmatrix} \times \Phi(\mathbb{X})$; alors

$$\Psi(0) = \begin{pmatrix} 1 & 0 \\ -\alpha & 1 \end{pmatrix} \times \Phi(0) = \begin{pmatrix} 1 & 0 \\ -\alpha & 1 \end{pmatrix}$$

(car $\Phi(0) = \mathcal{I}d$), donc $\Psi(0)$ est bien inversible et le modèle est correctement spécifié.

Montrer que la régression \mathcal{R} par les moindres carrés de X_t sur $\langle X_{t-1}, \dots, X_{t-p} \rangle$ est équivalente aux deux estimations séparées \mathcal{R}^1 de X_t^1 sur $\langle X_{t-1}^1, X_{t-1}^2, \dots, X_{t-p}^1, X_{t-p}^2 \rangle$ et \mathcal{R}^2 de X_t^2 sur $\langle X_t^1, X_{t-1}^1, X_{t-1}^2, \dots, X_{t-p}^1, X_{t-p}^2 \rangle$.

D'une façon générale, l'estimateur des moindres carrés ordinaires de la régression de $Y = Xb + U$ est, si X est de plein rang, $\hat{b}_{mcg} = (X'\Omega^{-1}X)^{-1}X' \times \Omega^{-1}Y$, où Ω désigne la matrice de variance-covariance de U .

Considérons alors les deux sous-blocs

$$\begin{pmatrix} Y^1 \\ Y^2 \end{pmatrix} = \begin{pmatrix} Xb^1 \\ Xb^2 \end{pmatrix} + \begin{pmatrix} U^1 \\ U^2 \end{pmatrix}$$

et supposons que $\mathbb{V} \begin{pmatrix} U^1 \\ U^2 \end{pmatrix} = \begin{pmatrix} \Omega^{1,1} & (0) \\ (0) & \Omega^{2,2} \end{pmatrix}$ soit bloc-diagonale.

Alors

$$\begin{aligned} \hat{b}_{mcg} &= (X'\Omega^{-1}X)^{-1}X' \times \Omega^{-1} \begin{pmatrix} Y^1 \\ Y^2 \end{pmatrix} \\ &= \left(X' \begin{pmatrix} \Omega^{1,1} & (0) \\ (0) & \Omega^{2,2} \end{pmatrix}^{-1} X \right)^{-1} X' \times \begin{pmatrix} \Omega^{1,1} & (0) \\ (0) & \Omega^{2,2} \end{pmatrix}^{-1} \begin{pmatrix} Y^1 \\ Y^2 \end{pmatrix} \\ &= \begin{pmatrix} (X'(\Omega^{1,1})^{-1}X)^{-1}X' \times (\Omega^{1,1})^{-1}Y^1 \\ (X'(\Omega^{2,2})^{-1}X)^{-1}X' \times (\Omega^{2,2})^{-1}Y^2 \end{pmatrix} \\ &= \begin{pmatrix} \hat{b}_{mcg}^1 \\ \hat{b}_{mcg}^2 \end{pmatrix} \end{aligned}$$

Dans le cas présent, la régression \mathcal{R} s'écrit

$$\begin{pmatrix} 1 & 0 \\ -\alpha & 1 \end{pmatrix} \begin{pmatrix} X_t^1 \\ X_t^2 \end{pmatrix} = \begin{pmatrix} X_{t-1}^1 & \dots & X_{t-p}^1 & X_{t-1}^2 & \dots & X_{t-p}^2 \end{pmatrix} \times \begin{pmatrix} \psi_1^{1,1} & \psi_1^{2,1} \\ \vdots & \vdots \\ \psi_p^{1,1} & \psi_p^{2,1} \\ \psi_1^{1,2} & \psi_1^{2,2} \\ \vdots & \vdots \\ \psi_p^{1,2} & \psi_p^{2,2} \end{pmatrix} + \begin{pmatrix} \eta_t^1 \\ \eta_t^2 \end{pmatrix}$$

tandis que les sous-régressions \mathcal{R}^1 et \mathcal{R}^2 s'écrivent respectivement

$$(\mathcal{R}^1): X_t^1 = \begin{pmatrix} X_{t-1}^1 & \dots & X_{t-p}^1 & X_{t-1}^2 & \dots & X_{t-p}^2 \end{pmatrix} \times \begin{pmatrix} \psi_1^{1,1} \\ \vdots \\ \psi_p^{1,1} \\ \psi_1^{1,2} \\ \vdots \\ \psi_p^{1,2} \end{pmatrix} + \eta_t^1$$

et

$$(\mathcal{R}^2): X_t^2 = \begin{pmatrix} X_t^1 & X_{t-1}^1 & \dots & X_{t-p}^1 & X_{t-1}^2 & \dots & X_{t-p}^2 \end{pmatrix} \times \begin{pmatrix} \alpha \\ \psi_1^{2,1} \\ \vdots \\ \psi_p^{2,1} \\ \psi_1^{2,2} \\ \vdots \\ \psi_p^{2,2} \end{pmatrix} + \eta_t^2$$

comme $\mathbb{V}(\eta_t) = \begin{pmatrix} \Sigma^{1,1} & 0 \\ 0 & \Sigma^{2,2} - \frac{(\Sigma^{1,2})^2}{\Sigma^{1,1}} \end{pmatrix}$ est diagonale, l'estimation des coefficients de Ψ par la régression globale \mathcal{R} est équivalente à celle résultant des deux sous-régressions \mathcal{R}^1 et \mathcal{R}^2 .

☞ Q3

On considère l'hypothèse \mathcal{H}_0 : " $X^1 \not\sim X^2$ ".

Réécrire les régressions \mathcal{R}^1 et \mathcal{R}^2 sous \mathcal{H}_0 .

Montrer qu'à supposer qu' ϵ^1 est gaussien, le test de l'hypothèse \mathcal{H}_0 se ramène à un test de Student que l'on explicitera.

L'hypothèse \mathcal{H}_0 se réécrit " $\Sigma^{1,2} = 0$ " soit encore " $\alpha = 0$ ".

Remarquons en outre que si ϵ est gaussien, alors il en va de même pour $\eta = \begin{pmatrix} \epsilon^1 \\ \epsilon^2 - \frac{\text{Cov}(\epsilon^2, \epsilon^1)}{\text{V}(\epsilon^1)} \epsilon^1 \end{pmatrix}$;

tester \mathcal{H}_0 revient donc à tester la nullité individuelle du coefficient estimé $\hat{\alpha}$, ce qui peut être fait au moyen de la statistique de Student habituelle.

Remarquons que sous \mathcal{H}_0 seule la régression \mathcal{R}^2 est modifiée; il suffit donc de pratiquer le test dans la seule régression \mathcal{R}^2 .

☞ Q4

On considère l'hypothèse \mathcal{H}_0^* : " $X^2 \not\sim X^1$ ".

En supposant qu' ϵ est gaussien, proposer une procédure de test de \mathcal{H}_0^* .

L'hypothèse \mathcal{H}_0^* se réécrit " $\forall k \in \mathbb{N}, \phi_k^{1,2} = 0$ " soit encore " $\forall k \in \mathbb{N}, \psi_k^{1,2} = 0$ ".

Le test de \mathcal{H}_0^* revient alors au test de la nullité simultanée des coefficients estimés $(\widehat{\psi}_1^{1,2}, \dots, \widehat{\psi}_p^{1,2})$, qui peut être conduit au moyen de la statistique de Fisher.¹¹

Remarquons que cette fois sous \mathcal{H}_0^* seule la régression \mathcal{R}^1 est modifiée, et tester \mathcal{H}_0^* revient à tester la régression \mathcal{R}^1 sous \mathcal{H}_0^* par rapport à \mathcal{R}^1 sous $\neg\mathcal{H}_0^*$. Ainsi il suffit de pratiquer le test dans la seule régression \mathcal{R}^1 .

¹¹ A savoir $F = \frac{SCR_{\mathcal{H}_0^*} - SCR_{\neg\mathcal{H}_0^*}}{SCR_{\neg\mathcal{H}_0^*}} \cdot \frac{T-2p}{p}$ où T désigne le nombre d'observations, $2p$ le nombre total de variable explicatives et p le nombre de variables dont on teste la nullité simultanée, et bien-sûr $SCR_{\mathcal{H}_t}$ désigne la somme des carrés des résidus estimés (i.e. la variance non-expliquée) sous l'hypothèse \mathcal{H} . Si \mathcal{H}_0 est vraie alors F suit une loi de Fisher $\mathcal{F}(p, T - 2p)$.