ENSAE - 2ème année

Séries temporelles linéaires
Examen 2018-2019

Durée : 2 heures. Sans document.

Les exercices sont indépendants. Il est demandé de justifier les réponses de façon concise.

Exercice 1 Soit $(u_t)_{t\in\mathbb{N}}$ une suite de variables aléatoires réelles indépendantes et identiquement distribuées telles que $Eu_t = m$ et $\mathrm{Var}(u_t) = \sigma^2$ existent. On définit la suite $(X_t)_{t\in\mathbb{N}}$ par $X_0 = 0$ et $X_t = \rho X_{t-1} + u_t$ pour $t \geq 1$.

1. La suite $(u_t)_{t\in\mathbb{N}}$ est-elle toujours ergodique ? A quelle condition est-elle un bruit blanc ? Quelle est la limite presque sûre de $\frac{1}{n}\sum_{t=1}^{n} u_t^2$ quand $n \to \infty$?

   2 pts : La suite $(u_t)_{t\in\mathbb{N}}$ étant iid, elle est strictement stationnaire et ergodique. C'est un bruit blance si $m = 0$. Par le théorème ergodique, la limite presque sûre de $\frac{1}{n}\sum_{t=1}^{n} u_t^2$ quand $n \to \infty$ est $Eu_t^2 = \sigma^2 + m^2$.

2. On note $\mu_t = EX_t$. Exprimer μ_t en fonction de ρ, m et μ_{t-1}, puis en fonction de ρ, m et t.

   2 pts : On a $\mu_1 = m$ et pour $t \geq 2$

   $$\mu_t = m + \rho\mu_{t-1} = m + \rho m + \rho^2\mu_{t-2} = m\left\{1 + \rho + \rho^2 + \cdots \rho^{t-1}\right\}$$

   .

3. La figure 1 représente une trajectoire $X_0, X_1, \ldots, X_n$ de longueur $n = 300$. A la vue de ce graphique, pouvez-vous rejeter l'hypothèse que $m = 0$? Pouvez-vous rejeter l'hypothèse que $\rho = 1$? Est-il nécessaire de faire des tests statistiques pour cela ?

   2 pts : Si on avait $m = 0$ on aurait $\mu_t = 0$ pour tout t. Si on a $|\rho| < 1$, cela devrait se voir sur la trajectoire car on a dans ce cas un phénomène de retour à la moyenne. La trajectoire de la figure n'est visiblement pas centrée, donc on peut rejeter l'hypothèse que $m = 0$ (et même deviner

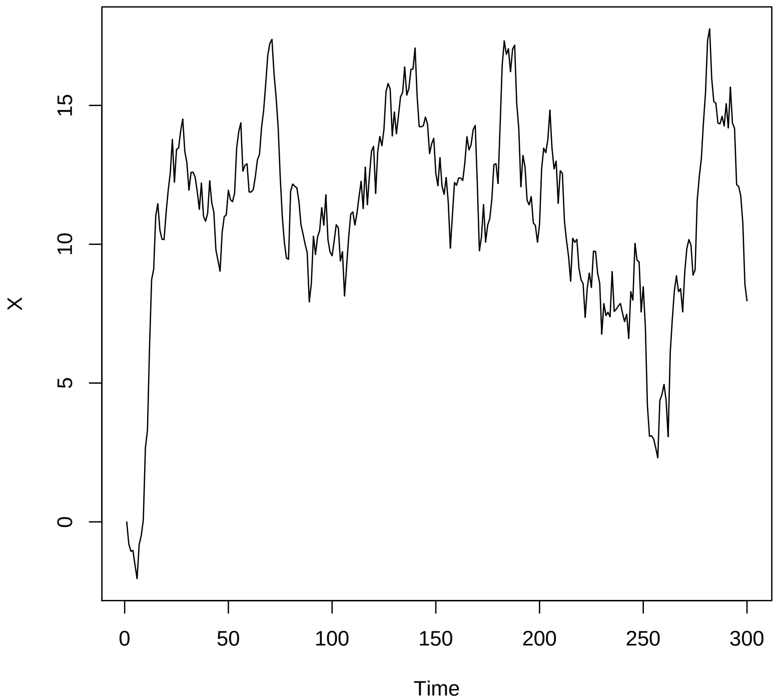

FIGURE 1 – Trajectoire de la série (X_t).

que $m > 0$) dans le cas $|\rho| < 1$. Si on avait $\rho = 1$ avec $m \neq 0$ la trajectoire devrait posséder la tendance déterministe $t \mapsto mt$, ce qui n'est manifestement pas le cas, donc on peut rejeter l'hypothèse que $\rho = 1$ quand $m \neq 0$. Il est inutile de faire un test statistique pour une telle conclusion (si on admet le modèle AR(1) avec constante de l'énoncé, on se trouve dans le cas 3 du cours sur les tests de racine unité). Le seul cas non trivial à voir sur la trajectoire et $\rho = 1$ et $m = 0$. On peut donc envisager le test de Dickey-Fuller où H_0 est l'hypothèse d'une marche aléatoire sans dérive contre l'alternative H_1 d'un AR(1) avec $|\rho| < 1$ et m éventuellement non nul (cas 2 du cours).

Exercice 2 Soit $(\epsilon_{1t}, \epsilon_{2t}, \epsilon_{3t})'$ un bruit blanc fort de variance identité I_3, et $X_t = (X_{1t}, X_{2t}, X_{3t})'$ satisfaisant

$$\begin{cases} X_{1t} &= aX_{3t} + \epsilon_{1t} \\ X_{2t} &= bX_{1t} + \epsilon_{2t} \\ X_{3t} &= cX_{3,\,t-1} + \epsilon_{3t}. \end{cases}$$

1. On suppose dans cette question que $|c| < 1$.

(a) Ecrire ce système sous forme VAR(1) en précisant la variance du bruit et montrer qu'il satisfait la condition d'existence d'une solution stationnaire et non anticipative.

2 pts : Puisque $X_{1t} = acX_{3,t-1} + a\epsilon_{3t} + \epsilon_{1t}$ et $X_{2t} = bacX_{3,t-1} + ba\epsilon_{3t} + b\epsilon_{1t} + \epsilon_{2t}$, on a $X_t = AX_{t-1} + \epsilon_t$ avec

$$A = \begin{pmatrix} 0 & 0 & ac \\ 0 & 0 & bac \\ 0 & 0 & c \end{pmatrix}, \quad \epsilon_t = \begin{pmatrix} a\epsilon_{3t} + \epsilon_{1t} \\ ba\epsilon_{3t} + b\epsilon_{1t} + \epsilon_{2t} \\ \epsilon_{3t} \end{pmatrix}.$$

La variance de ϵ_t est

$$\Sigma = \begin{pmatrix} 1+a^2 & b(1+a^2) & a \\ b(1+a^2) & 1+b^2+a^2b^2 & ab \\ a & ab & 1 \end{pmatrix}.$$

Le polynôme $\det(I_3 - Az) = 1 - cz$ n'a pas de racine à l'intérieur ou sur le cercle unité, donc le modèle VAR(1) admet une solution stationnaire non anticipative.

(b) Le vecteur $(X_{1t}, X_{2t})'$ cause-t-il X_{3t} au sens de Granger ? Pour quelles valeurs de a, b et c la variable X_{3t} cause-t-elle le vecteur $(X_{1t}, X_{2t})'$ au sens de Granger ? A-t-on causalité instantanée entre X_{1t} et $(X_{2t}, X_{3t})'$?

2 pts : D'après l'emplacement des 0 dans la matrice A, le vecteur $(X_{1,t-1}, X_{2,t-1})'$ n'intervient pas dans l'expression X_{3t}, et donc $(X_{1t}, X_{2t})'$ ne cause pas X_{3t} au sens de Granger. Pour la même raison, X_{3t} cause le vecteur $(X_{2t}, X_{3t})'$ au sens de Granger si $a \neq 0$ et $c \neq 0$. Daprès la forme de Σ, il y a causalité instantanée entre X_{1t} et $(X_{2t}, X_{3t})'$ si le vecteur $(b(1+a^2), a)$ est non nul, donc si $a \neq 0$ ou $b \neq 0$.

2. On suppose dans cette question que $c = 1$. Ecrire ce système sous forme à correction d'erreur VECM. Pour quelles valeurs de a et b le processus X_t est-il cointégré ? Quel est son rang de cointégration ?

2 pts : La troisième équation est une marche aléatoire et les deux premières équations sont des relations de cointégration. Si $a \neq 0$ ou $b \neq 0$ le système est donc cointégré au sens de Granger (il est cointégré au sens large pour toutes valeurs de a et b). Le rang de cointégration est

2. On a la représentation VECM de forme $\nabla X_t = \Pi X_{t-1} + \epsilon_t$ avec

$$\Pi = \begin{pmatrix} -1 & 0 & ac \\ 0 & -1 & abc \\ 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} -1 & 0 \\ -b & -1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & -a \\ b & 1 & 0 \end{pmatrix}.$$

On retrouve les relations de cointégration dans cette décomposition, ainsi que le rang de cointégration.

Exercice 3 Il arrive souvent que l'on veuille prévoir une séries temporelle Y_t en fonction de ses valeurs passées et également d'une variable X_t dite "exogène". Ceci peut se faire à l'aide de modèles appelés ARMAX, dont nous considérons dans cet exercice la version la plus simple.

Soit (ϵ_t) un bruit blanc et (X_t) une série temporelle univariée telle que X_t soit observable avant la variable d'intérêt Y_t. On suppose que la série bivariée (X_t, ϵ_t) est strictement stationnaire et ergodique, avec $EX_t^2 < \infty$, et on considère le modèle AR(1)-X

$$Y_t = aY_{t-1} + bX_t + c + \epsilon_t, \quad t \in \mathbb{Z}.$$

1. On suppose $|a| < 1$, mais on n'impose aucune contrainte sur b et c. Donner la solution Y_t stationnaire et ergodique de l'équation AR(1)-X. On suppose que ϵ_t est indépendant de $\{X_u, u \leq t; \epsilon_u, u < t\}$. Quelle est la meilleure prévision de Y_t en fonction de $\{X_u, u \leq t; Y_u, u < t\}$? Quelle est la variance du terme d'erreur ?

   2 pts : En itérant, on voit que la solution stationnaire est

   $$Y_t = \sum_{i=0}^{\infty} a^i (bX_{t-i} + c + \epsilon_{t-i}).$$

   Cette série est bien absolument convergente avec probabilité 1 car

   $$E \sum_{i=0}^{\infty} |a|^i |bX_{t-i} + c + \epsilon_{t-i}| \leq \frac{|b|E|X_1| + |c| + E|\epsilon_1|}{1 - |a|} < \infty.$$

   Au sens des moindres carré, la prévision optimale est

   $$\hat{Y}_t = aY_{t-1} + bX_t + c$$

   avec le terme d'erreur ϵ_t de variance σ_ϵ^2

2. On étudie dans cette question les conséquences de l'oubli de la variable exogène. On suppose que $X_t = \eta_t + d\eta_{t-1}$ avec $|d| < 1$ et (η_t) un bruit blanc fort de variance σ_η^2, indépendant du bruit fort (ϵ_t) de variance σ_ϵ^2. Posons $Z_t = bX_t + \epsilon_t$.
   (a) Déterminer la fonction d'autocovariance de (Z_t). Quel est le modèle ARMA suivi par Z_t ? Quel est le modèle ARMA suivi par Y_t ?

   2 pts : On a

   $$\mathrm{Var}(Z_t) = \sigma_\epsilon^2 + b^2\sigma_\eta^2(1 + d^2), \quad \mathrm{Cov}(Z_t, Z_{t-1}) = b^2 d\sigma_\eta^2$$

   et $\mathrm{Cov}(Z_t, Z_{t-h}) = 0$ pour $h > 1$, donc $Z_t \sim MA(1)$ de la forme $Z_t = u_t + \varphi u_{t-1}$ où u_t est un bruit blanc faible de variance σ_u^2. Par conséquent $Y_t \sim ARMA(1,1)$.

   (b) Quelle est approximativement la variance du terme d'erreur du modèle ARMA suivi par (Y_t) quand $|b|$ est très grand ? Quelle est alors la conséquence de l'oubli de la variable exogène pour la prévision de Y_t ?

   2 pts : Les paramètres φ et σ_u^2 de la MA(1) sont tels que

   $$\gamma_Z(0) = \sigma_u^2(1 + \varphi^2) = \sigma_\epsilon^2 + b^2\sigma_\eta^2(1 + d^2),$$
   $$\gamma_Z(1) = \sigma_u^2\varphi = b^2\sigma_\eta^2 d.$$

   Lorsque σ_ϵ^2 est négligeable devant $b^2\sigma_\eta^2$, par identification , on trouve

   $$\varphi \sim d, \qquad \sigma_u^2 \sim b^2\sigma_\eta^2.$$

   Puisque $\sigma_u^2 >> \sigma_\epsilon^2$, l'oubli de la variable exogène dégrade fortement la qualité de la prévision.

3. On étudie dans cette question les conséquences de l'oubli de la dynamique de Y_t. Pour que les calculs soient simples, on suppose maintenant que la suite (X_t) est iid indépendante de (ϵ_t). Dans le modèle de régression

   $$Y_t = \tilde{b}X_t + \tilde{c} + e_t,$$

   où e_t est centré et non corrrélé avec X_t, que valent $\tilde{b}$, $\tilde{c}$ et la variance de e_t ? Quelle est alors la conséquence de l'oubli de la dynamique de Y_t ? Pour estimer les paramètres de ce modèle, peut-on faire confiance aux sorties des logiciels de régression usuels ?

2 pts : Avec des notations évidentes, on a

$$\mathrm{Var}(Y_t) = \frac{b^2\sigma_X^2 + \sigma_\epsilon^2}{1-a^2}, \qquad EY_t = \frac{b\,m_X + c}{1-a}, \quad \mathrm{Cov}(X_t, Y_t) = b\sigma_X^2$$

d'où

$$\tilde{b} = \frac{\mathrm{Cov}(X_t, Y_t)}{\mathrm{Var}(X_t)} = b, \qquad \tilde{c} = EY_1 - \tilde{b}EX_1 = \frac{c + ab\,m_X}{1-a}$$

et

$$\sigma_e^2 = \mathrm{Var}Y_t - \frac{\{\mathrm{Cov}(X_t, Y_t)\}^2}{\mathrm{Var}(X_t)} = \frac{b^2a^2\sigma_X^2 + \sigma_\epsilon^2}{1-a^2}.$$

Puisque $\sigma_e^2 > \sigma_\epsilon^2$ on perd également en négligeant la dynamique. Comme le couple (X_t, Y_t) est stationnaire ergodique, les estimateurs MCO de la régression convergent, mais les lois asymptotiques (t-statistiques en particulier) ne sont pas les mêmes que celles d'une régression ordinaire, ce qui rend délicat l'interprétation des sorties des logiciels standard.