

Les Processus saisonniers

Gilbert Colletaz

21 juillet 2020

Résumé

On peut souvent observer dans les trajectoires de variables économiques des phénomènes qui se répètent régulièrement au cours du temps et cela pendant une durée constante. Ces caractéristiques sont révélatrices de la présence d'une saisonnalité qui affecte la variable étudiée. La durée constante définit la période de la saisonnalité, typiquement l'année pour des observations mensuelles ou trimestrielles, la semaine pour des séries journalières, une journée pour des séries horaires. Bien évidemment la modélisation de ce type de séries nécessite la prise en compte de ces saisonnalités et donc l'adaptation des processus étudiés jusqu'à présent. C'est la raison d'être des processus SARMA(p,q) × (P,Q)_s. Après une discussion générale sur la nécessité de modéliser les phénomènes saisonniers et une présentation intuitive de ces nouveaux processus, nous étudierons les filtres les plus simples de cette classe afin de voir quelles sont les évolutions attendues de leurs fonctions d'autocorrélation, *i.e.* comment se comportent les outils d'identification de la procédure de Box-Jenkins.

de

Table des matières

1	Les processus SARMA(p,q) × (P,Q)_s	2
1.1	la nécessaire prise en compte de la saisonnalité	2
1.2	la logique de construction des SARMA(p,q) × (P,Q) _s	2
2	Études de quelques processus saisonniers	3
2.1	le SARMA(0,0) × (0,1) ₄	3
2.1.1	ses fonctions d'autocovariance et d'autocorrélation :	4
2.1.2	sa fonction d'autocorrélation partielle	4
2.1.3	illustration avec $x_t = u_t - 0.8u_{t-4}$	4
2.2	Le SARMA(0,1) × (0,1) ₁₂	4
2.2.1	ses fonctions d'autocovariance et d'autocorrélation partielle	5
2.2.2	sa fonction de corrélation partielle	5
2.2.3	un exemple	5
2.2.4	remarque	5
2.3	Le SMA(1) × AR(1) ₇	6
2.3.1	ses fonctions d'autocovariance et d'autocorrélation	6
2.3.2	sa fonction de corrélation partielle	7
2.3.3	un exemple	7
3	Exemple de traitement : le modèle <i>airline</i> de Box-Jenkins	8
4	Conclusion	8
4.1	résumé des évolutions des fonctions d'autocorrélations	8
4.2	l'estimation des SARMA(p,q) × ARMA(P,Q) _s	9
4.2.1	avec la Proc ARIMA sous SAS	9
4.2.2	avec la commande ARIMA de R	9

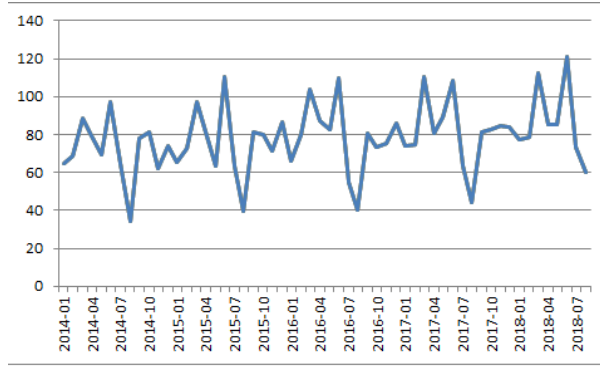


FIGURE 1 – Immatriculations neuves - voitures particulières - marques françaises (en milliers)

1 Les processus SARMA(p,q) × (P,Q)_s

1.1 la nécessaire prise en compte de la saisonnalité

La figure 1 qui représente les immatriculations de voitures neuves de marques françaises sur la période 2014-2018 illustre les caractéristiques d’une série saisonnière. Ici les observations sont mensuelles et on remarque bien au cours de chaque année des événements récurrents. Par exemple, des pics sur les nombres des immatriculations sont observés aux mois de mars et juin, et un creux tous les mois d’août, et des profils relativement similaires sur les mois du dernier trimestre des différentes années : la saisonnalité a ici une période de 12 mois.

Ce simple graphique montre bien que les modèles vus jusqu’ici doivent être amendés pour tenir compte de ces phénomènes saisonniers. En effet en règle générale dans ces modèles non saisonniers, les observations les plus similaires à celle observée à une date t sont celles des dates immédiatement antérieures : $x_{t-1}, x_{t-2}, \dots, x_{t-k}$, où k est souvent relativement faible, ce qui se traduit par une décroissance assez rapide des autocorrélations. En revanche si on considère ici les immatriculations du mois de mars ou du mois d’août d’une année quelconque, on remarque que le nombre qui en est le plus proche est celui du mois de mars ou du mois d’août des années précédentes, et non pas celles des mois de février ou juillet qui les précèdent immédiatement. En d’autres termes, sur cette série mensuelle, les corrélations les plus élevées ne seront souvent pas les premières, mais plutôt celles observées au voisinage de $\rho_{12}, \rho_{24}, \rho_{36}, \dots$, *i.e.* au voisinage des corrélations dont les ordres sont des multiples de la longueur de la saison, le *span*, égal à douze dans cet exemple.

Ainsi dans la liste des processus en mesure de modéliser une trajectoire on va ajouter aux MA(q), AR(p), ARMA(p,q) des filtres saisonniers, les SARMA(p,q) × (P,Q)_s, où s est la période de la saisonnalité que l’on veut modéliser, et, comme nous allons le voir, p et q sont les ordres des composantes ARMA non saisonnières, et P et Q ceux des composantes ARMA saisonnières

1.2 la logique de construction des SARMA(p,q) × (P,Q)_s

Pour comprendre ces nouveaux filtres, il peut être utile de revenir sur ce que sont les modèles MA, AR ou ARMA non saisonniers. On peut parler d’eux comme des filtres dans la mesure où à partir d’observations constituant la trajectoire d’une variable quelconque, observations généralement linéairement dépendantes, ils restituent, après application de polynômes en L une série exempte d’autocorrélations, soit schématiquement :

variable initiale non saisonnière	filtre polynômes en L	résidu autocorrélations nulles
$x_t \rightarrow$	MA(q)	↘
	ou AR(p)	→ u_t
	ou ARMA(p,q)	↗

Maintenant avec les séries saisonnières, on conçoit qu’il existe deux types de dépendances au sein de la séquence des observations. D’une part des dépendances non saisonnières, identiques à celles prises en compte dans le schéma qui précède, traduisant le fait que l’observation d’une date t dépend des observations ou/et des innovations qui lui sont antérieures mais relativement proches sur l’échelle du temps. D’autre part des dépendances sur des observations et/ou des innovations qui lui sont proches au regard de la saisonnalité mais qui, sur l’échelle du temps se trouvent à des dates dont l’antériorité

est au voisinage de multiples du span.

Dans ces conditions, si on passe la série saisonnière dans un des filtres du schéma précédent, filtres constitués de polynômes en L , et si ces filtres continuent de remplir leur fonction, on va récupérer à la sortie une série y qui aura été épurée de ses dépendances non saisonnières mais dans laquelle subsistera les dépendances saisonnières. Ce filtre, semblable à ceux vus jusqu'ici, correspond à la partie ARMA(p,q) du SARMA(p,q) \times (P,Q)_s. Il suffit alors pour retrouver un résidu exempt de toute autocorrélation, saisonnière ou non, de passer la série y dans un filtre adapté qui va supprimer les dépendances saisonnières. Pour cela, il faudra écrire y_t non pas en fonction de y_{t-1}, y_{t-2}, \dots mais en fonction de y_{t-s}, y_{t-2s}, \dots , i.e. les polynômes de ce filtre auront comme argument non pas L mais L^s , et ils correspondront à la composante ARMA(P,Q) du SARMA(p,q) \times (P,Q)_s.

Au final, l'enchaînement qui conduit d'une série x possédant des dépendances saisonnières et non saisonnières à une série u ou l'une et l'autre sont absentes s'effectue en deux étapes :

première étape		
variable initiale	filtre polynômes en L	série exempte de dépendance non saisonnière
$x_t \rightarrow$	MA(q) ou AR(p) ou ARMA(p,q)	\searrow $\rightarrow y_t$ \nearrow
deuxième étape		
série en entrée	filtre polynômes en L^s	série exempte de toute dépendance saisonnière et non saisonnière
$y_t \rightarrow$	MA(Q) ou AR(P) ou ARMA(P,Q)	\searrow $\rightarrow u_t$ \nearrow

On peut enfin aisément reconstruire l'équation du SARMA(p,q) \times (P,Q)_s, c'est-à-dire du filtre qui prend x en entrée et renvoie u en sortie. Il suffit pour cela de recombinaison les deux filtres non saisonniers (1) et saisonniers (2) :

$$\phi(L)x_t = \theta(L)y_t \quad (1)$$

$$\Phi(L^s)y_t = \Theta(L^s)u_t \quad (2)$$

$$(3)$$

Sous l'hypothèse d'un $\theta(L)$ inversible, on a avec (1), $y_t = \theta(L)^{-1}\phi(L)x_t$, et cette valeur de y_t remise dans (2) donne finalement l'équation du SARMA(p,q) \times (P,Q)_s :

$$\phi(L)\Phi(L^s)x_t = \theta(L)\Theta(L^s)u_t \quad (4)$$

où :

- $\phi(L)$ est la composante autorégressive non saisonnière,
- $\Phi(L^s)$ est la composante autorégressive saisonnière,
- $\theta(L)$ est la composante moyenne mobile non saisonnière,
- $\Theta(L^s)$ est la composante moyenne mobile saisonnière.

et le processus d'innovation u_t sera causal pour x_t si les deux polynômes autorégressifs sont inversibles de sorte à retrouver l'écriture de Wold sur la variable stationnaire x . Comme d'habitude, cela va exiger que les racines de ces polynômes soient, en module, supérieures à l'unité.

2 Études de quelques processus saisonniers

2.1 le SARMA(0, 0) \times (0, 1)₄

Ce processus ne comporte aucune composante non saisonnière et sa composante saisonnière est de type MA. La période étant de 4, il sera plutôt adapté à des observations trimestrielles. Son équation est :

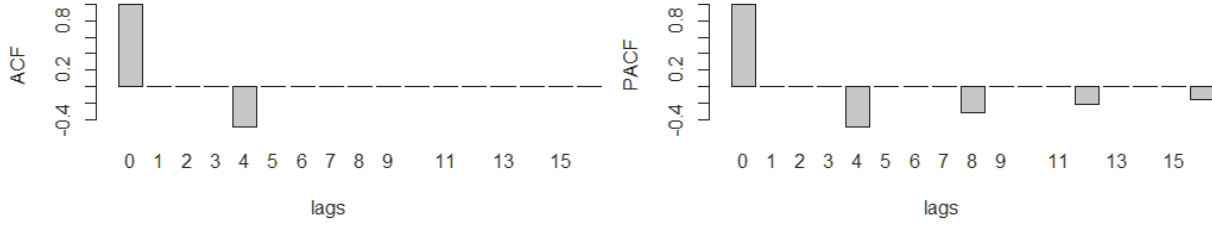


FIGURE 2 – Autocorrélations du SMA(1) : $x_t = u_t - 0.8u_{t-1}$

$$\begin{aligned} X_t &= \mu_X + (1 - \Theta_1 L^4)u_t, \text{ soit avec } x_t = X_t - \mu_X : \\ x_t &= u_t - \Theta_1 u_{t-4} \end{aligned} \quad (5)$$

On peut le regarder comme un MA(4) troué, il est donc stationnaire. Par ailleurs, le polynôme caractéristique de ce SMA possède 4 racines, $-i\sqrt[4]{\frac{1}{\Theta_1}}$, $i\sqrt[4]{\frac{1}{\Theta_1}}$, $-\sqrt[4]{\frac{1}{\Theta_1}}$, $\sqrt[4]{\frac{1}{\Theta_1}}$, qui pour être hors du cercle complexe unitaire exigent $|\Theta_1| < 1$. Sous cette condition, x_t possède une écriture autorégressive infinie :

$$\begin{aligned} (1 - \Theta_1 L^4)^{-1}x_t &= u_t, \text{ soit encore} \\ x_t &= -\sum_{j=1}^{\infty} \Theta_1^j x_{t-4j} + u_t = -\Theta_1 x_{t-4} - \Theta_1^2 x_{t-8} - \Theta_1^3 x_{t-12} - \dots + u_t \end{aligned}$$

2.1.1 ses fonctions d'autocovariance et d'autocorrélation :

On vous laisse vérifier que :

- $\gamma_0 = E[x_t^2] = (1 + \Theta_1^2) \sigma_u^2$,
- $\gamma_4 = E[x_t x_{t-4}] = -\Theta_1 \sigma_u^2$,
- $\gamma_j = E[x_t x_{t-j}] = 0$ si $j \neq 0, 4$.

en conséquence, $\rho_0 = 1$, $\rho_4 = -\Theta_1/(1 + \Theta_1^2)$ et $\rho_j = 0$ pour $j \notin \{0, 4\}$. On retrouve notamment un résultat connu : ce MA saisonnier est un MA non saisonnier troué de longueur 4 dont la fonction d'autocorrélation doit s'annuler au-delà de l'ordre 4.

2.1.2 sa fonction d'autocorrélation partielle

Comme c'est un AR(∞), elle ne s'annulera jamais au-delà d'un certain retard. Avec ce processus saisonnier simple, on peut montrer aisément que seules les autocorrélations partielles d'ordres égaux à des multiples du span sont non nulles et décroissent régulièrement.

2.1.3 illustration avec $x_t = u_t - 0.8u_{t-4}$

On rappelle qu'avec la convention de signe, le coefficient de ce processus est $\Theta_1 = 0.8 > 0$ et donc $\rho_4 = -0.488$. La Figure 2 présente ses fonctions d'autocorrélation et d'autocorrélation partielle.

2.2 Le SARMA(0, 1) \times (0, 1)₁₂

Une variable gouvernée par ce processus possède donc des dépendances non saisonnières et des dépendances saisonnières qui sont éliminées par deux filtres de type moyenne mobile. Son équation est :

$$X_t = \mu_X + (1 - \theta_1 L)(1 - \Theta_1 L^{12})u_t, \text{ soit avec } x_t = X_t - \mu_X : \quad (6)$$

$$x_t = (1 - \theta_1 L)(1 - \Theta_1 L^{12})u_t \text{ ou encore,}$$

$$x_t = u_t - \theta_1 u_{t-1} - \Theta_1 u_{t-12} + \theta_1 \Theta_1 u_{t-13} \quad (7)$$

C'est un MA(13) trouvé, il est stationnaire. Pour posséder une écriture autorégressive équivalente, qui sera alors infinie, il faut que les deux polynômes MA soient inversibles. Le premier, non saisonnier, exige $|\theta_1| < 1$, et le second, que chacune de ses 12 racines soit extérieure au cercle complexe unitaire.

2.2.1 ses fonctions d'autocovariance et d'autocorrélation partielle

On sait déjà que ses autocovariances d'ordre supérieures à 13 seront nulles. Vous devez ensuite pouvoir repérer que seules 5 autocovariances sont non nulles et que leurs valeurs sont :

- $\gamma_0 = E[x_t^2] = (1 + \theta_1^2 + \Theta_1^2 + \theta_1^2 \Theta_1^2) \sigma_u^2 = (1 + \theta_1^2) (1 + \Theta_1^2) \sigma_u^2$
- $\gamma_1 = E[x_t x_{t-1}] = -(\theta_1 + \theta_1 \Theta_1^2) \sigma_u^2 = -\theta_1 (1 + \Theta_1^2) \sigma_u^2$
- $\gamma_{11} = E[x_t x_{t-11}] = (\theta_1 \Theta_1) \sigma_u^2$
- $\gamma_{12} = E[x_t x_{t-12}] = -(\Theta_1 + \theta_1^2 \Theta_1) \sigma_u^2 = -\Theta_1 (1 + \theta_1^2) \sigma_u^2$
- $\gamma_{13} = E[x_t x_{t-13}] = (\theta_1 \Theta_1) \sigma_u^2$

En conséquence,

- $\rho_0 = 1$,
- $\rho_1 = -\frac{\theta_1}{1 + \theta_1^2}$
- $\rho_{11} = \rho_{13} = \frac{\theta_1 \Theta_1}{(1 + \theta_1^2)(1 + \Theta_1^2)}$
- $\rho_{12} = -\frac{\Theta_1}{1 + \Theta_1^2}$
- $\rho_j = 0$ si $j \notin \{0, 1, 11, 12, 13\}$

2.2.2 sa fonction de corrélation partielle

Sous les habituelles conditions de racines des deux polynômes MA, le processus a une écriture AR(∞) et cette fonction ne s'annule jamais. Elle va décroître vers zéro, avec des valeurs qui vont toutefois avoir tendance à réaugmenter en valeur absolue au voisinage des multiples du span.

2.2.3 un exemple

On illustre les divers points précédents en simulant une trajectoire donnée par : $x_t = (1 - 0.6L)(1 - 0.8L^{12})u_t$, où u est un bruit blanc gaussien de variance unitaire. Les autocorrélations non nulles sont alors : $\rho_0 = 1$, $\rho_1 = -0.441$, $\rho_{11} = \rho_{13} = 0.215$ et $\rho_{12} = -0.488$.

La figure 3 présente les fonctions d'autocorrélations et d'autocorrélations partielles estimées au moyen d'une trajectoire simulée de 600 observations. Vous pouvez vérifier qu'avec ces observations les corrélations estimées, r_1 , r_{11} , r_{12} et r_{13} ont les signes attendus et sont significatives. On repère également des autocorrélations partielles estimées qui décroissent, avec des pics aux retards 1, 12, 24, 36 et 48, et des réaugmentations à leurs voisinages. On vous invite toutefois à reproduire cet exercice sur des séries de longueurs différentes à celle employée ici afin de percevoir la dépendance de ces évolutions attendues au nombre d'observations. Notez enfin que le choix d'un MA saisonnier est renforcé par l'examen de ces graphiques : les autocorrélations pointent vers un SMA(1)×MA(1) avec les 4 corrélations significatives, et donc avec un processus ayant, hors constante et variance résiduelle, deux coefficients à estimer. En revanche, avec 4 autocorrélations significatives dans les cinq premières et celles qui réapparaissent au voisinage des multiples du span égal à 12, un AR saisonnier exigerait l'estimation d'un nombre nettement plus important de coefficients.

2.2.4 remarque

Avec le SMA(0,1) × (0,1) qui précède, on a montré que la corrélation d'ordre 12, correspondant à la longueur du span, est encadrée de deux corrélations égales aux ordres 11 et 13. Ceci est un cas particulier d'un énoncé plus général : dans un modèle saisonnier de période s de type SMA(0,q)×(0,Q), la corrélation d'ordre s est encadrée de q corrélations symétriques, *i.e.* telles que $\rho_{s-1} = \rho_{s+1}$, $\rho_{s-2} = \rho_{s+2}$, ..., $\rho_{s-q} = \rho_{s+q}$: la longueur de la symétrie est égale à la longueur de la composante MA non saisonnière. Vérifiez ainsi que sur un SARMA(0,2)×(0,1)₁₂ on a bien $\rho_{10} = \rho_{14} \neq 0$ et $\rho_{11} = \rho_{13} \neq 0$.

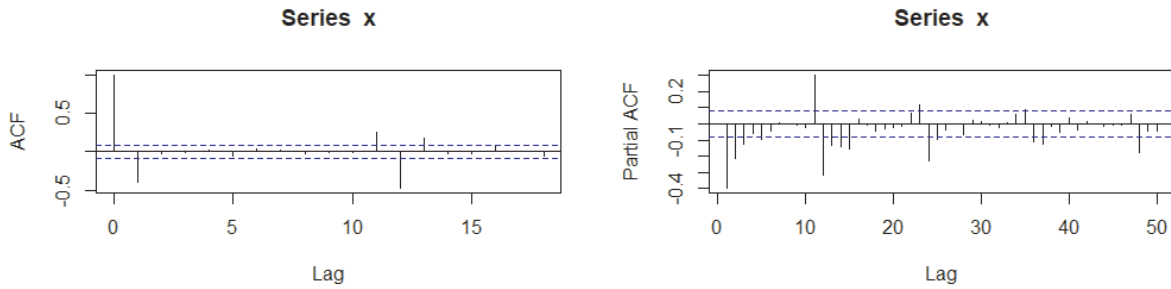


FIGURE 3 – Autocorrélations du SMA(1) : $x_t = u_t - 0.8u_{t-1}$

2.3 Le SMA(1) \times AR(1)₇

Dans les exemples qui précèdent, nous n'avons considéré que des polynômes MA. Ici nous allons mixer une composante AR saisonnière et une composante MA non saisonnière, les deux de longueur 1 afin de rester sur des calculs très simples. Si une variable x centrée est gouvernée par ce processus, alors :

$$(1 - \Phi_1 L^7)x_t = (1 - \theta_1 L)u_t, \text{ ou encore :} \quad (8)$$

$$x_t = \Phi_1 x_{t-7} + u_t - \theta_1 u_{t-1} \quad (9)$$

On sait alors que la stationnarité du processus n'est pas assurée, et vous devez savoir qu'elle exige $|\Phi_1| < 1$ afin que les 7 racines du polynôme autorégressif soient toutes de module supérieur à l'unité. Dans ce cas, le processus possède une écriture équivalente dans la classe des MA(∞). Par ailleurs, si $|\theta_1| < 1$ alors le polynôme MA(1) est inversible et il existe également un processus qui lui est équivalent dans la classe des AR(∞).

2.3.1 ses fonctions d'autocovariance et d'autocorrélation

x_t s'écrit comme une somme d'éléments orthogonaux entre eux et si on ajoute l'hypothèse de stationnarité, il vient immédiatement que sa variance est $\gamma_0 = \frac{1+\theta_1^2}{1-\Phi_1^2} \sigma_u^2$. Par ailleurs,

$$\gamma_1 = E[x_t x_{t-1}] = \Phi_1 E[x_{t-7} x_{t-1}] + \underset{=0}{E[u_t x_{t-1}]} - \theta_1 E[u_{t-1} x_{t-1}]$$

et

$$E[u_{t-1} x_{t-1}] = \Phi_1 E[x_{t-8} u_{t-1}] + \underset{=0}{E[u_{t-1}^2]} - \theta_1 \underset{=0}{E[u_{t-2} u_{t-1}]} = \sigma_u^2$$

D'où $\gamma_1 = \Phi_1 \gamma_6 - \theta_1 \sigma_u^2$.

A partir de l'ordre 2 on a dépassé la mémoire de la composante MA, et donc :

$$\gamma_j = E[x_t x_{t-j}] = \Phi_1 E[x_{t-7} x_{t-j}] = \Phi_1 \gamma_{j-7} \text{ pour } j \geq 2$$

Ce dernier résultat donne en particulier :

$$\gamma_k = \Phi_1^j \gamma_0 \text{ pour } k = 7j \text{ et } j = 0, 1, 2, \dots$$

mais aussi $\gamma_6 = \gamma_8 = \Phi_1 \gamma_1$, $\gamma_{13} = \gamma_{15} = \Phi_1^2 \gamma_1$ et plus généralement :

$$\gamma_k = \Phi_1^j \gamma_1 \text{ pour } k = 7j \pm 1 \text{ et } j = 1, 2, \dots$$

ce qui permet d'obtenir $\gamma_1 = -\frac{\theta_1 \sigma_u^2}{1-\Phi_1^2}$, soit encore $\rho_1 = \gamma_1 / \gamma_0 = -\frac{\theta_1}{1+\theta_1^2}$.

Enfin, pour les valeurs de k non considérées ci-dessus, on a d'une part $\gamma_k = \Phi_1 \gamma_{k-7}$ et d'autre part, en raison de la symétrie de la fonction d'autocovariance $\gamma_{k-7} = \gamma_{7-k}$. Comme $\gamma_{7-k} = \Phi_1 \gamma_{-k} = \Phi_1 \gamma_k$, il faut donc vérifier $\gamma_k = \Phi_1^2 \gamma_k$. La seule solution possible lorsque $\Phi_1 \neq 0$ et sous hypothèse de stationnarité est naturellement $\gamma_k = 0$.

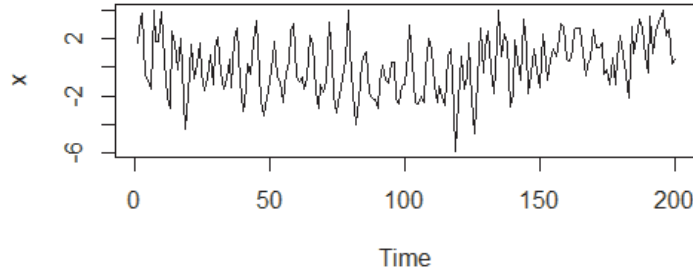


FIGURE 4 – Trajectoire simulée du $\text{SMA}(1) \times \text{AR}(1)_7 : x_t = 0.8_1 x_{t-7} + u_t - 0.6u_{t-1}$

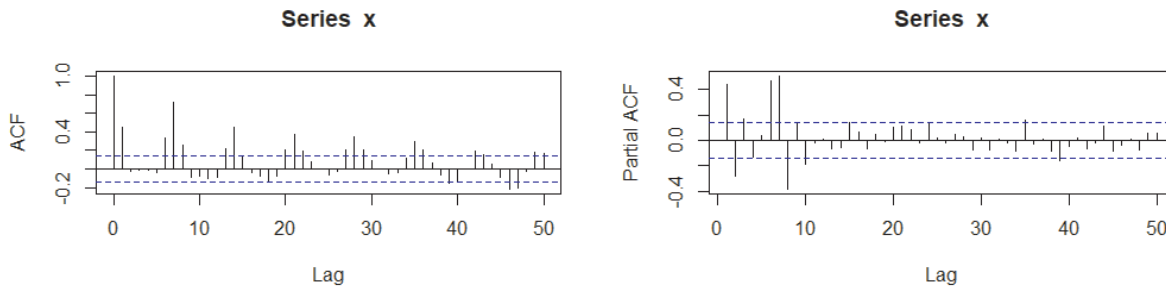


FIGURE 5 – Autocorrélations du $\text{SMA}(1) \times \text{AR}(1)_7 : x_t = 0.8_1 x_{t-7} + u_t - 0.6u_{t-1}$

Au final, les autocorrélations de ce processus $\text{SMA}(1) \times \text{AR}(1)_7$ sont données par :

$$\rho(k) = \begin{cases} 1 & \text{si } k = 0, \\ \Phi_1^j & \text{si } k = 7j, j = 0, 1, 2, 3, \dots, \\ -\frac{\theta_1}{1+\theta_1^2} \Phi_1^j & \text{si } k = 7j \pm 1, j = 0, 1, 2, \dots, \\ 0 & \text{dans tous les autres cas.} \end{cases}$$

On observerait une décroissance exponentielle des autocorrélations aux multiples du span et un encadrement symétrique de chacune d'elles par un nombre de corrélations égal à la longueur de la composante MA non saisonnière, ici de longueur unitaire.

2.3.2 sa fonction de corrélation partielle

On aura une décroissance irrégulière des coefficients d'autocorrélation partielle avec notamment des pics aux multiples du span.

2.3.3 un exemple

On a simulé sur 200 points le processus $x_t = 0.8_1 x_{t-7} + u_t - 0.6u_{t-1}$. Les figures 4 et 5 présentent respectivement la trajectoire obtenue et les autocorrélations estimées qui lui correspondent. Avec cet exemple, on repère aisément la décroissance attendue des corrélations aux multiples du span, ce qui signale la présence d'une composante autorégressive saisonnière, et leur encadrement par une seule corrélation significative indiquant la présence d'une composante MA d'ordre 1. Par ailleurs, aucune régularité n'est mise en évidence sur les corrélations partielles d'ordres égaux aux multiples du span, ce qui tend aussi à invalider la présence d'un MA saisonnier.

3 Exemple de traitement : le modèle *airline* de Box-Jenkins

Comme dernière illustration, nous reprenons en cours l'exemple de la modélisation du nombre de passagers des vols aériens aux US. Les données sont mensuelles et observées de janvier 1949 à décembre 1960. Dans leur ouvrage Box et Jenkins modélisent cette série au moyen d'un processus $SIMA(1) \times IMA(1)_{12}$, *i.e.* selon $\Delta\Delta^{12}x_t = (1 - \theta_1 L)(1 - \Theta_1 L^{12})$. Ce modèle est certainement aujourd'hui le processus saisonnier le plus utilisé en pratique, il est connu sous le nom de *airline model*.

Un simple graphique de la série en question montre que son niveau tend à augmenter dans le temps, qu'il existe une saisonnalité de span 12 et que l'amplitude des fluctuations croît également dans le temps. On peut conclure alors évidemment que cette série ne respecte absolument pas l'hypothèse de stationnarité exigée par la modélisation ARMA ou SARMA. Nous verrons que la transformation proposée pour stabiliser la variance est de prendre le logarithme des valeurs initiales : on ne va pas modéliser x_t mais $y_t = \log(x_t)$, et vous pouvez constater visuellement que l'opération remplit son office : les amplitudes des fluctuations de y_t semblent pratiquement constantes au contraire de ce que l'on observait sur x_t . La seule question que l'on traite ici est de savoir pourquoi le passage au log parvient ici à cette stabilisation de la variance de la série transformée.

Soit donc y_t une transformée de x_t par une fonction continue dérivable $f(\cdot)$. Un développement de Taylor autour d'une valeur x_0 donne :

$$y_t = f(x_t) \approx f(x_0) + f'(x_0)(x_t - x_0) \text{ et donc,}$$

$$Var(y_t) \approx [f'(x_0)]^2 Var(x_t)$$

On se rappelle qu'avec la série brute, l'amplitude des fluctuations et le niveau des observations augmentent tous deux avec le temps. Ceci est compatible avec l'hypothèse d'une variance de x qui serait proportionnelle au niveau de x . Il suffirait alors de trouver une fonction $f(x_t)$ dont la dérivée serait inversement proportionnelle à x_t pour tendre vers le résultat recherché. On peut penser alors à la fonction $\log(x_t)$ et à sa dérivée $1/x_t$. Tout se cale parfaitement si la variance de x_t n'est pas proportionnelle à son niveau mais au carré de celui-ci puisqu'en effet, si $Var(x_t) = kx_t^2$ avec k une constante positive, alors

$$y_t = \log(x_t) \Rightarrow Var(y_t) \approx \left(\frac{1}{x_t}\right)^2 kx_t^2 = k, \tag{10}$$

et la variance de la transformée par passage aux logarithmes est bien une constante. Naturellement pour parvenir à ce résultat, il faut que l'hypothèse $Var(x_t) = kx_t^2$ soit raisonnable, comme cela semble être le cas sur la série de Box et Jenkins. Vous pouvez par exemple vérifier que si la variance de x_t est proportionnelle non pas au carré de son niveau, mais seulement à son niveau, alors la transformée y_t aurait une variance et donc des fluctuations qui devraient tendre à se réduire dans le temps. Vous pouvez vérifier qu'alors la transformation adaptée serait de définir $y_t = x_t^{1/2}$ en supposant évidemment $x_t > 0$.

4 Conclusion

4.1 résumé des évolutions des fonctions d'autocorrélations

On peut résumer dans la table 1 le comportement attendu des fonctions usuelles d'identification des processus dans la démarche de Box-Jenkins pour des modèles saisonniers purs, *i.e.* sans composante non saisonnière. On retrouve des schémas déjà vus pour les processus non saisonniers, ils sont simplement maintenant appliqués aux autocorrélations d'ordres égaux à des multiples du span caractérisant la saisonnalité de la série étudiée.

processus de span s	autocorrélations $\rho(k)$	autocorrélations partielles ϕ_{kk}
SMA(Q)	$\rho(k) = 0$ pour $k > sQ$	$\lim_{k \rightarrow \infty} \phi_{sk, sk} = 0$
SAR(P)	$\lim_{k \rightarrow \infty} \rho_{sk} = 0$	$\phi_{kk} = 0$ pour $k > sP$
SARMA(P,Q)	$\lim_{k \rightarrow \infty} \rho_k = 0$	$\lim_{k \rightarrow \infty} \phi_{k, k} = 0$

TABLE 1 – Évolution des autocorrélations et autocorrélations partielles pour des purs processus saisonniers

Naturellement la présence de composantes non saisonnières complique sensiblement l'étape d'identification en faisant surgir d'autres corrélations significatives autour de celles précisées dans cette table. L'expérience du statisticien est sans

doute ici un élément extrêmement favorable, mais notez toutefois qu'il est toujours possible de recourir aux critères de sélection du type Akaike et/ou Schwarz lorsqu'a été définie une famille de modèles concurrents. Il est ainsi possible de sortir de cette étape d'identification fondée sur l'examen des corrélations en ayant retenu plusieurs processus, pour laisser un des critères sélectionner le filtre optimal qui devra également être soumis aux tests usuels de validation.

4.2 l'estimation des SARMA(p,q) × ARMA(P,Q)_s

4.2.1 avec la Proc ARIMA sous SAS

L'indication des puissances de l'opérateur de retard L présentes dans les polynômes MA ou AR s'effectue toujours via les options `p=` et/ou `q=`. La présence de polynômes saisonniers est indiquée selon une syntaxe qui mime l'écriture des SARMA(p,q) × ARMA(P,Q)_s : dans celle-ci les polynômes saisonniers et non saisonniers AR d'une part et MA d'autre part se multiplient entre eux selon : $\phi(L)\Phi(L^s)(X_t - \mu_X) = \theta(L)\Theta(L^s)u_t$. On va faire de même pour les listes parenthésées des puissances de L correspondant à chacun de ces polynômes. Ainsi

- pour réclamer l'estimation de $(1 - \phi_1 L - \phi_2 L^2)(1 - \phi_1 L^{12})(X_t - \mu_X) = (1 - \theta_1 L)(1 - \Theta_2 L^{12} - \Theta_3 L^{24})$ on fera apparaître dans la commande `estimate` : `p=(1,2)(12) q=(1)(12,24)` ;
- le SMA(1) × MA(1)₇ sera estimé avec `q=(1)(7)`.

On a toujours le choix de la procédure d'estimation via l'option `method=CLS/u1s/ml`, et la possibilité d'éliminer la constante du modèle avec `noconstant` ou `noint`.

Notez que dans cette procédure on n'a pas besoin de spécifier le span de la saisonnalité prise en compte puisque les puissances de L doivent être explicitement indiquées.

4.2.2 avec la commande ARIMA de R

On retrouve toujours la commande `arima`, simplement en plus du vecteur associé à `order = c(p,0,q)` spécifiant les ordres des composantes AR et MA non saisonnières, on indique les composantes saisonnières via : `seasonal = list(order = c(P, 0, Q), period = s)`.

Remarquez que contrairement à SAS, on doit avec cette formulation spécifier la période de la saisonnalité, s . Enfin, la présence ou non d'un terme constant dans l'équation est toujours gouvernée par la présence de `include.mean=T/F`.

- Remarque : la commande `arma.sim` n'est pas adaptée à la simulation de processus ARMA saisonniers. Si vous désirez vous livrer à ce type d'exercice, utilisez le package `gmwm` qui remplit cette fonction très simplement. Par exemple pour obtenir 200 observations du processus SARMA(1,1) × (2,1)₁₂, $(1 - 0.3L)(1 - 0.12L^{12} - 0.20L^{24})x_t = (1 - 0.27L)(1 - 0.9L^{12})u_t$, où u est un bruit blanc gaussien d'écart-type 1.5, il vous suffira de faire :

```
sarima=gen.gts(SARIMA(ar=0.3, i=0, ma=-0.27, sar=c(-0.12, -0.2), si=0, sma =-0.9, sigma2=1.5, s=12), 200)
```

Ce package peut être installé via les commandes :

```
# install.packages("devtools")
devtools::install_github("smac-group/gmwm")
```