L'identification chez Box-Jenkins

Gilbert Colletaz

1er octobre 2020

Résumé

On connaît maintenant les propriétés des fonctions d'autocorrélation, d'autocorrélation partielle et inverse pour chacune des trois classes de processus. On sait également qu'au moins pour deux d'entres elles, MA(q) et AR(p), il n'est pas impossible que l'examen de ces fonctions permette d'identifier la classe d'appartenance et, au sein de la classe, l'ordre du processus qui leur correspond. La difficulté est qu'en pratique on ne connaît pas les vraies valeurs des autocorrélations mais seulement leurs estimations. Il s'agit donc maintenant de savoir comment s'obtiennent les estimateurs de ces autocorrélations, quelles sont leurs propriétés et comment on vont-elles être les utilisées dans la procédure de recherche du filtre générateur d'une série d'observations.

Table des matières

1 Les autocorrélations estimées

1.1 La fonction d'autocorrélation

On sait que si on note γ_k la covariance de x_t et de x_{t-k} alors la corrélation entre ces deux variables est par définition :

$$\begin{aligned}\rho_k &= \frac{cov(x_t, x_{t-k})}{var(x_t)^{1/2} var(x_{t-k})^{1/2}}, \text{ soit encore sous hypothèse de stationnarité} \\ &= \frac{\gamma_k}{\gamma_0^{1/2}\gamma_0^{1/2}} \\ &= \frac{\gamma_k}{\gamma_0}, \; k = 1, 2, 3, \ldots \end{aligned} \tag{1}$$

Sur un échantillon d'observations $x_1, x_2 \ldots, x_T$, l'estimateur du maximum de vraisemblance de γ_k est donné par

$$\hat{\gamma}_k = \frac{1}{T}\sum_{t=1}^{T-k}(x_{t+k} - \bar{x})(x_t - \bar{x}), \; k = 1, 2, 3, \ldots \tag{2}$$

où $\bar{x}$ est la moyenne empirique, $\bar{x} = \frac{\sum_{t=1}^{T} x_t}{T}$.

Soit $r_k = \frac{\hat{\gamma}_k}{\hat{\gamma}_0} = \frac{\sum_{t=1}^{T-k}(x_{t+k}-\bar{x})(x_t-\bar{x})}{\sum_{t=1}^{T}(x_t-\bar{x})^2}$, comme $plim\, \hat{\gamma}_k = \gamma_k$, on a

$$plim\, r_k = plim\, \frac{\hat{\gamma}_k}{\hat{\gamma}_0} = \frac{plim\, \hat{\gamma}_k}{plim\, \hat{\gamma}_0} = \frac{\gamma_k}{\gamma_0} = \rho_k \tag{3}$$

Cette corrélation empirique est donc un estimateur convergent de la corrélation théorique. L'expression de la variance asymptotique de cet estimateur donnée par Bartlett est relativement complexe :

$$var(r_k) = \frac{1}{T} \sum_{j=-\infty}^{+\infty} (\rho_j^2 + \rho_{k+j}\rho_{-k+j} - 4\rho_k\rho_j\rho_{-k+j} + 2\rho_j^2\rho_k^2) \tag{4}$$

Ainsi, la précision de r_k dépend d'une infinité de corrélations théoriques qu'évidemment on ne connaît généralement pas. Il faudra donc réaliser un tour de passe-passe pour que l'équation ci-dessus nous permette d'évaluer la variance de r_k.

1.2 La fonction d'autocorrélation partielle

On sait que la corrélation partielle d'ordre k, ϕ_{kk} apparaît dans le processus autorégressif d'ordre k d'écriture :

$$x_t = c + \phi_{k1}x_{t-1} + \phi_{k2}x_{t-2} + \ldots + \phi_{k,k-1}x_{t-k+1} + \phi_{kk}x_{t-k} + v_t \tag{5}$$

Les estimations de ces coefficients d'autocorrélation partielle peuvent donc être obtenues par des ajustements de type OLS : $\hat{\phi}_{kk}$ est donné par le coefficient de x_{t-k} dans la régression de x_t sur $x_{t-1}, x_{t-2}, \ldots, x_{t-k}$. Une autre possibilité est de prendre les équations de Yule-Walker dans lesquelles les autocorrélations théoriques, ρ_k, sont remplacés par les estimations r_k obtenues comme vu au point précédent. Enfin Quenouille a montré que l'écart-type de ces estimateurs $\hat{\phi}_{kk}$ a une expression plus simple que celle des autocorrélations puisqu'il est simplement égal à T^{-1}.

2 La détermination de l'ordre d'un AR au moyen des autocorrélations partielles estimées

Dans cette première section, on suppose être en présence d'un AR d'ordre inconnu. Si on pouvait observer les ϕ_{kk}, on trouverait aisément cet ordre p, puisque :

$$p = \inf\{k|\phi_{kk} = 0\}$$

La difficulté vient de ce que sur l'échantillon d'observations disponibles on a été seulement en mesure de calculer $\hat{\phi}_{kk}$, l'estimateur du maximum de vraisemblance de ϕ_{kk}. Celui-là est un "bon" estimateur au sens où $E[\hat{\phi}_{kk}] = \phi_{kk}$, et $plim(\hat{\phi}_{kk}]) = \phi_{kk}$, mais il reste que $Pr[\hat{\phi}_{kk} = \phi_{kk}] = 0$: même si les vraies valeurs des autocorrélations partielles sont nulles, les estimations ne le seront pas. Il est donc nécessaire de tester la nullité des coefficients inconnus puisque l'ordre à partir duquel on ne rejettera pas leur nullité sera naturellement l'ordre AR estimé du processus.
La construction d'un test ponctuel ne pose alors pas de difficulté majeure puisque l'on connaît parfaitement la distribution asymptotique de $\hat{\phi}_{kk}$: les estimateurs du maximum de vraisemblance sont asymptotiquement gaussiens, ils sont sans biais, et on connaît leur leur variance : $1/T$.

Dans ces conditions,

$$\hat{\phi}_{kk} \underset{T\to\infty}{\xrightarrow{d}} \mathcal{N}(\phi_{kk}, 1/T)$$

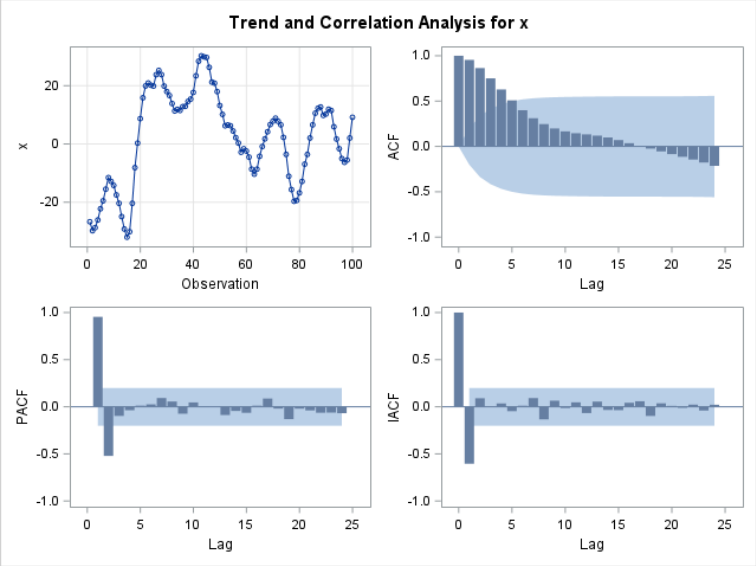
et un intervalle de confiance à 95% construit autour de la valeur testée, $\phi_{kk} = 0$ est simplement égal à $\pm 2/\sqrt{T}$: toute valeur estimée qui sort de cet intervalle permet de rejeter la nullité du vrai coefficient d'autocorrélation partielle au seuil de risque de 5%. Il s'en suit une séquence de questions et de tests :

1. est-on en présence au moins d'un AR(1) ?
   Le test associé est H0 :$\phi_{11} = 0$ versus H1 :$\phi_{11} \neq 0$.
   La règle de décision : si $|\hat{\phi}_{11}| > 2/\sqrt{T}$ alors rejet de H0, sinon, non rejet de H0.
   La conclusion : si non rejet, c'est un AR d'ordre inférieur à 1, i.e. c'est un bruit blanc (**rappel, on ne traite pour l'instant que des processus *non troués***). Si rejet, on est en présence au moins d'un AR(1), ce qui conduit à la question suivante :
2. est-on en présence au moins d'un AR(2) ?
   Le test associé est H0 :$\phi_{22} = 0$ versus H1 :$\phi_{22} \neq 0$.
   La règle de décision : si $|\hat{\phi}_{22}| > 2/\sqrt{T}$ alors rejet de H0, sinon, non rejet de H0.
   La conclusion : si non rejet, c'est un AR d'ordre inférieur à 2, i.e. c'est un AR(1) . Si rejet, on est en présence au moins d'un AR(2), ce qui conduit à la question suivante :

3. est-on en présence au moins d'un AR(3) ?
   Le test associé est H0 :$\phi_{33} = 0$ versus H1 :$\phi_{33} \neq 0$.
   La règle de décision : si $|\hat{\phi}_{33}| > 2/\sqrt{T}$ alors rejet de H0, sinon, non rejet de H0.
   La conclusion : si non rejet, c'est un AR d'ordre inférieur à 3, i.e. c'est un AR(2) Si rejet, on est en présence au moins d'un AR(3), ce qui conduit à la question suivante :
4. est-on en présence au moins d'un AR(4) ?
   ...

L'enchaînement continue jusqu'à ce que l'hypothèse nulle ne soit pas rejetée et on a alors l'ordre estimé de l'AR.

En pratique la plupart des logiciels donnent en sortie les corrélogrammes estimés avec indication de l'intervalle de confiance, souvent à 95%. La décision finale peut ainsi apparaître rapidement comme le montre l'exemple suivant pour lequel nous avons utilisé les graphes de la Proc ARIMA de SAS.



Dans le premier quadrant on va trouver la série des observations, ici l'échantillon est de 100 valeurs. Dans le deuxième on trouve les 24 premiers coefficients de la fonction d'autocorrélation estimée laquelle se caractérise par une décroissance régulière, caractéristique soit d'un AR, soit d'un ARMA. Dans le troisième on a la fonction d'autocorrélation partielle estimée dont l'évolution est plus marquée par une cassure que par une décroissance régulière, ce qui est caractéristique d'un AR et non pas d'un MA ou d'un ARMA. La fonction d'autocorrélation inverse du quatrième quadrant est marquée également par une cassure plutôt que par une décroissance régulière, *i.e.* elle est représentative des autocorrélations d'un MA. Tous ces éléments vont dans le même sens : le choix du modèle qui a généré les observations du premier quadrant va plutôt pencher vers un processus AR. Il reste à trouver l'ordre de ce dernier.
Avant de se livrer à cet exercice, il faut noter que dans cette procédure SAS, le graphe des corrélations et des corrélations inverses commence au rang $k = 0$, alors qu'il débute au rang $k = 1$ pour les partielles. Enfin, la bande colorée en bleu clair représente l'intervalle de confiance à 95% construit autour de zéro. On peut d'ailleurs voir que dans cette procédure ARIMA, la variance des autocorrélations inverses est égale à celle des autocorrélations partielles, *i.e.* $1/T$.

L'observation des partielles signale alors que $\hat{\phi}_{11}$ et $\hat{\phi}_{22}$ entraîne respectivement le rejet de H0 :$\phi_{11} = 0$ et H0 :$\phi_{22} = 0$. En revanche, pour $k = 3, 4, 5 \ldots$ on ne rejette pas la nullité des ϕ_{kk}. Ces corrélations partielles feraient donc retenir un AR(2). Si on se tourne vers les corrélations inverses, hormis naturellement celle de rang zéro qui est égale à 1, seule celle de rang 1 serait non nulle ce qui caractérise les corrélations d'un MA(1) et donc ce dernier graphique ferait retenir un AR(1) sur la série étudiée. Au final, à la fin de cette étape d'identification, les deux processus AR(1) ou AR(2) peuvent être raisonnablement sélectionnés et dans ce cas le choix final sera délégué à l'étape d'estimation et aux tests de validation d'un processus estimé.

3 La détermination de l'ordre d'un MA au moyen des autocorrélations estimées

On suppose maintenant que l'on sait être en présence d'un MA et qu'il reste à trouver l'ordre q de ce MA. On sait que cet ordre est repéré sur les corrélations selon :

$$q = \inf\{k|\, \rho_k = 0\}$$

La démarche employée sera dans sa logique exactement la même que celle qui vient d'être présenté pour la détermination du choix de l'ordre d'un AR. La seule différence vient du calcul de la variance des estimateurs r_k : la formule de Bartlett est en effet sensiblement plus complexe que celle de Quenouille qui s'appliquait aux $\hat{\phi}_{kk}$. En reprenant l'équation de Bartlett, et en se souvenant qu'on suppose uniquement des MA non troués, vous devez pouvoir montrer que :

$Var(r_1) = \frac{1}{T}$si toutes les autocorrélations sont nulles à l'exception évidemment de $\rho_0, i.e.$ on est en présence d'un $MA(0)$

$Var(r_2) = \frac{1}{T}(1 + 2\rho_1^2) = Var(r_1) + \frac{2\rho_1^2}{T}$, si seules ρ_0 et ρ_1 sont non nulles, $i.e$ on est en présence d'un $MA(1)$

$Var(r_3) = \frac{1}{T}(1 + 2\rho_1^2 + 2\rho_2^2) = Var(r_2) + \frac{2\rho_2^2}{T}$, si seules ρ_0, ρ_1 et ρ_2 sont non nulles, $i.e$ on est en présence d'un $MA(2)$

$\vdots$

$Var(r_k) = \frac{1}{T}(1 + 2\sum_{j=1}^{k-1} \rho_i^2) = Var(r_{k-1}) + \frac{2\rho_{k-1}^2}{T}$, si seules $\rho_0, \rho_1, \ldots, \rho_{k-1}$ sont non nulles, $i.e$ on est en présence d'un MA(k-1)

On déjà tirer deux enseignements de ces équations :

- d'une part on voit que la variance de r_k augmente avec k, ainsi la largeur d'un intervalle de confiance construit sur r_k pour un seuil de risque donné va augmenter avec k. Cependant cette augmentation va avoir tendance à se ralentir puisque lorsque k augmente, ρ_k va avoir tendance à diminuer.
- d'autre part, on note qu'en pratique elles sont inutilisables pour le calcul de la variance de l'estimateur puisqu'elles sont fonction des vraies valeurs de corrélations, ρ_k, valeurs qui sont a priori inconnues. Pour sortir de cette impasse, la solution usuelle est simplement de remplacer ces vraies valeurs par leurs estimations r_k. On a alors :

$$\begin{aligned}
\text{pour le MA(0)}, \widehat{Var}(r_1) &= \frac{1}{T} \\
\text{pour le MA(1)}, \widehat{Var}(r_2) &= \frac{1}{T}(1 + 2r_1^2) = \widehat{Var}(r_1) + \frac{2r_1^2}{T} \\
\text{pour le MA(2)}, \widehat{Var}(r_3) &= \frac{1}{T}(1 + 2r_1^2 + 2r_2^2) = \widehat{Var}(r_2) + \frac{2r_2^2}{T} \\
&\vdots \\
\text{pour le MA(k-1)}, \widehat{Var}(r_k) &= \frac{1}{T}(1 + 2\sum_{j=1}^{k-1} r_i^2) = \widehat{Var}(r_{k-1}) + \frac{2r_{k-1}^2}{T}
\end{aligned}$$

Comme $plim\, r_k = \rho_k$, il vient $plim\, \widehat{Var}(r_k) = Var(r_k)$: on a des estimateurs convergents des variances. On note également toujours l'augmentation de la variance estimée avec l'ordre de l'autocorrélation de sorte on observera une augmentation de la taille des intervalles de confiance construits sur les r_k avec k et cela notamment sur les faibles valeurs de k, là où souvent se situent les corrélations empiriques les plus élevées.

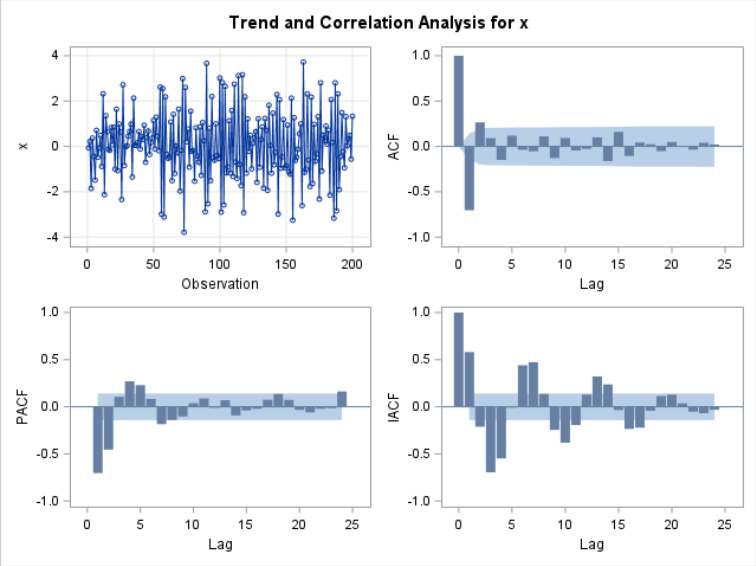
La recherche de l'ordre du MA se fera alors selon un enchaînement maintenant connu de questions et de tests :

1. est-on en présence au moins d'un MA(1) ?
   Le test associé est H0 :$\rho_1 = 0$ versus H1 :$\rho_1 \neq 0$.
   La règle de décision : si $|r_1| > 2/\sqrt{T}$ alors rejet de H0, sinon, non rejet de H0.
   La conclusion : si non rejet, c'est un MA d'ordre inférieur à 1, i.e. c'est un bruit blanc et, si rejet, on est en présence au moins d'un MA(1), ce qui conduit à la question suivante :

2. est-on en présence au moins d'un MA(2) ?
   Le test associé est H0 : $\rho_2 = 0$ versus H1 : $\rho_2 \neq 0$.
   La règle de décision : si $|r_2| > 2\sqrt{\widehat{Var}(r_2)}$ alors rejet de H0, sinon, non rejet de H0.
   La conclusion : si non rejet, c'est un MA d'ordre inférieur à 2, i.e. c'est un MA(1) . Si rejet, on est en présence au moins d'un MA(2), ce qui conduit à la question suivante :
3. est-on en présence au moins d'un MA(3) ?
   Le test associé est H0 :$\rho_3 = 0$ versus H1 :$\rho_3 \neq 0$.
   La règle de décision : si $|r_3| > 2\sqrt{\widehat{Var}(r_3)}$ alors rejet de H0, sinon, non rejet de H0.
   La conclusion : si non rejet, c'est un MA d'ordre inférieur à 3, i.e. c'est un MA(2) Si rejet, on est en présence au moins d'un MA(3), ce qui conduit à la question suivante :
4. est-on en présence au moins d'un MA(4) ?
   ...

L'enchaînement continue jusqu'à ce que l'hypothèse nulle ne soit pas rejetée et on a alors l'ordre estimé du MA.

On illustre la démarche au moyen des graphes suivants toujours issus des sorties de la proc ARIMA de SAS.



Sur l'ACF, on repère plus une rupture qu'une décroissance ce qui ferait choisir un MA. Sur les autocorrélations partielles et sur les inverses on peut imaginer une décroissance sinusoïdale. On sait qu'une décroissance sur la PACF est un signal favorisant le choix d'un processus MA et que la décroissance sur IACF favorise un processus dual de type AR, et donc un MA sur la série étudiée. Les trois corrélogrammes ont donc ici des évolutions qui font retenir un MA.
L'ordre de celui-ci est repéré sur les autocorrélations : on voit facilement qu'au seuil de 5% l'hypothèse H0 : $\rho_1 = 0$ est rejetée, de même que H0 : $\rho_2 = 0$. En revanche on ne rejette pas H0 : $\rho_3 = 0$, H0 : $\rho_4 = 0$, et donc au final un MA(2) serait retenu. Au passage, vous pouvez noter l'effet d'élargissement de l'intervalle de confiance à 95% sur les r_k qui est surtout visible sur les deux-trois premièrs corrélations, ce qui est conforme au fait que $r1$ et r_2 sont les deux corrélations empiriques les plus grandes en valeur absolue et que toutes les autres sont proches de zéro.

Les deux exemples que nous venons de faire montrent l'utilité de la procédure d'identification de Box-Jenkins au moyen des fonctions de corrélation. Pour autant il ne faut pas en sous-estimer les limites :

— Si elle peut permettre de repérer l'adéquation d'un ARMA à une série, elle est pratiquement incapable de révéler les ordres de cet ARMA,
— si la décroissance d'une fonction s'effectue rapidement, on peut facilement confondre décroissance et annulation et donc se tromper sur la nature du processus. Par exemple, sur un AR(1) avec $\rho_1 = 0.3$, il vient $\rho_2 = 0.09$, $\rho_3 = 0.027$, $\rho_4 = 0.0081$. On peut concevoir alors que sur les estimateurs, cette parfaite décroissance des valeurs théoriques puisse ressembler à une rupture avec annulation.

— la procédure autorise une part non négligeable d'arbitraire : on peut imaginer que deux utilisateurs ayant une expérience différente de la méthode ne retiennent pas le même processus. A l'exception des évolutions évidentes, il est possible que l'un repère une décroissance là ou l'autre verra une rupture brutale vers zéro.
— enfin, elle est difficilement automatisable : utilisable s'il s'agit de sélectionner avec soin un modèle sur une série, on ne peut l'envisager si l'objectif est de prévoir des dizaines voire des centaines de séries en un temps relativement court.

Une solution pour pallier au moins au trois premiers inconvénients est de sortir de l'étape d'identification avec plusieurs modèles raisonnables, de les estimer et d'utiliser des tests de validation afin de choisir le processus final. Pour le dernier, il faut changer complètement la façon de procéder à l'identification et éliminer le plus possible l'intervention de l'utilisateur. C'est ce que vont permettre de faire les critères de sélection que nous étudierons par la suite.