

# Présentation et utilisation des critères de sélection

Gilbert Colletaz

12 septembre 2019

## Résumé

Dans cette partie nous allons présenter les critères de sélection d'Akaike, de Schwarz, et d'Hannan et Quinn qui sont les plus utilisés. Nous montrerons ensuite comment les mettre en application sous SAS et R pour rechercher un modèle optimal au sein d'un ensemble de modèles concurrents.

## Table des matières

<b>1</b>	<b>Présentation des principaux critères de sélection</b>	<b>1</b>
1.1	Le choix des modèles concurrents . . . . .	2
1.2	L'intuition pour la construction d'un critère . . . . .	3
1.3	Les critères de sélection les plus courants : Akaike, Schwarz, Hannan et Quinn . . . . .	4
1.3.1	probabilité de sélection d'un ordre inférieur à l'ordre vrai . . . . .	5
1.3.2	probabilité de sélection d'un ordre supérieur à l'ordre vrai . . . . .	5
1.4	Illustrations . . . . .	7
<b>2</b>	<b>Leur mise en oeuvre sous SAS et R</b>	<b>11</b>
2.1	L'option <code>minic</code> de SAS . . . . .	11
2.1.1	la démarche mise en oeuvre par <code>minic</code> . . . . .	11
2.1.2	un exemple . . . . .	11
2.2	Sous R : au moins deux commandes . . . . .	12
2.2.1	la commande <code>armaselect</code> du package <code>Caschrono</code> . . . . .	12
2.2.2	la commande <code>armasubsets</code> des packages <code>TSA</code> et <code>astsa</code> . . . . .	12
2.2.3	un exemple . . . . .	12

## 1 Présentation des principaux critères de sélection

L'emploi des fonctions de corrélation et corrélation partielle pour identifier un processus ARMA(p,q) implique une interprétation arbitraire de la part de l'utilisateur. Certes tout processus se caractérise par une représentation unique des valeurs théoriques de ces deux fonctions, dont la connaissance suppose d'ailleurs un certain degré d'expertise, mais leurs estimations décalent de ces valeurs théoriques, et cela d'autant plus que les tailles d'échantillon sont souvent relativement réduites. Ainsi, il n'est pas du tout improbable que plusieurs utilisateurs, auxquels seraient soumis les mêmes fonctions estimées, retiennent des filtres différents. On admet d'ailleurs que l'étape d'identification peut mener un seul utilisateur à retenir plusieurs modèles concurrents. C'est l'étape de validation qui, en principe, doit permettre de retenir parmi cet ensemble le processus "idéal" via des tests de significativité des coefficients, d'absence d'autocorrélation sérielle des résidus, ou bien encore via une confrontation out-of-sample de leurs capacités prédictives relatives.

Une autre façon d'aborder ce problème de sélection, qui lève une partie des difficultés précédentes, est d'imaginer que pour expliquer un phénomène économique on ait à sa disposition un ensemble de modèles concurrents et une règle de décision fondée sur l'emploi d'un critère de sélection. Si on est capable de calculer pour chacun de ces modèles la valeur du critère qui lui correspond, alors il suffit de retenir le filtre qui donne la meilleure valeur de ce critère. A l'évidence si plusieurs utilisateurs se mettent d'accord sur le critère en question d'une part et sur la famille de modèles concurrents d'autre part alors ils doivent normalement sélectionner le même filtre <sup>1</sup>.

Deux aspects sont donc à considérer :

---

1. aux options de programmation près

		p			
		0	1	...	$p_{max}$
q	0		AR(1)		AR( $p_{max}$ )
	1	MA(1)	ARMA(1,1)		ARMA( $p_{max},1$ )
	$\vdots$				
	$q_{max}$	MA( $q_{max}$ )	ARMA(1, $q_{max}$ )		ARMA( $p_{max},q_{max}$ )

TABLE 1 – Ensemble de modèles concurrents créé par  $(p_{max}, q_{max})$

- Le choix de l'ensemble des modèles concurrents
- Le choix du critère de sélection

Dans ce qui suit la présentation se fera pour l'essentiel dans le cadre d'une recherche des ordres  $p$  et  $q$  optimaux d'un processus ARMA stationnaire, sachant que l'utilisation de ces critères peut servir à aborder bien d'autres questions. Nous traiterons des deux points précédents, sachant qu'en ce qui concerne le second aspect, la dérivation analytiques des critères de sélection ne sera pas exposée. Nous donnerons seulement un aperçu de la logique présidant à l'emploi de ces critères pour ensuite en discuter trois, ceux d'Akaike <sup>2</sup>, de Schwarz <sup>3</sup> et de Hannan et Quinn <sup>4</sup> qui à eux seuls représentent pratiquement la totalité des critères employés dans la littérature.

## 1.1 Le choix des modèles concurrents

C'est l'aspect peut être le plus simple à résoudre. Pour commencer, supposons pour simplifier que la recherche se fasse parmi des processus ARMA non troués, ce qui par exemple exclut les modèles saisonniers. Il suffira alors de faire le pari que le vrai processus ARMA d'ordres  $p^*$  et  $q^*$  inconnus est tel que  $p^* < p_{max}$  et  $q^* < q_{max}$ . En d'autres termes, on postule que les ordres vrais sont respectivement inférieurs à deux ordres  $p_{max}$  et  $q_{max}$  que l'on se donne a priori. Le risque est ici naturellement de retenir des ordres  $p_{max}$  et  $q_{max}$  trop faibles qui n'encadrent pas les ordres vrais. En pratique l'examen des corrélogrammes représentatifs des autocorrélations et autocorrélations partielles estimées peut être utile pour fixer ces bornes maximales.

Le fait de fixer les bornes maximales  $p_{max}$  et  $q_{max}$  équivaut à définir une famille de  $(p_{max} + 1) \times (q_{max} + 1)$  filtres concurrents comme le montre le tableau 1.

C'est donc au sein de cette famille que va s'effectuer la recherche d'un modèle optimal au sens d'un certain critère. Notons déjà que si la procédure aboutit à la sélection d'un modèle appartenant à la dernière ligne ou à la dernière colonne alors la prudence commande d'imposer des valeurs pour  $p_{max}$  ou  $q_{max}$  supérieures à celles initialement choisies.

Ainsi que nous le verrons au point suivant, le nombre de paramètres à estimer est un élément important dans le calcul des différents critères de sélection. En conséquence la logique précédente ne peut pas s'appliquer sans ajustement dans le cas des processus ARMA saisonniers qui correspondent à des ARMA troués : par construction ils se caractérisent souvent par un ensemble non négligeable de coefficients nuls <sup>5</sup>. Dans ces conditions, si un phénomène saisonnier doit être décrit, il importe de construire la famille de modèles concurrents avec des modèles de type SARMA et non pas simplement en augmentant l'ordre des ARMA non troués <sup>6</sup>.

Par ailleurs, chaque modèle devant être estimé, il importe aussi de choisir des valeurs de  $p_{max}$  et  $q_{max}$  qui soient raisonnables au regard de la taille de l'échantillon servant aux estimations. Enfin il importe de s'assurer de la convergence des estimations notamment lorsque des composantes MA sont présentes : il arrive souvent que les algorithmes ne convergent pas lorsque trop de paramètres du type moyenne mobile sont impliqués <sup>7</sup>.

2. Akaike, H. (1974). New look at the statistical model identification. *IEEE Transactions on Automatic Control* **AC-19**, 716-723.

3. Schwarz, G. (1978). Estimating the dimension of a model. *Ann. Statist.* **6**, 461-464.

4. Hannan, E. J. and Quinn, B. G. (1979). The determination of the order of an autoregression, *J. Roy. Statist. Soc. Ser. B* **41**, 190-195.

5. Pour illustrer cette remarque et en anticipant sur leur traitement plus approfondi, supposez que sur données mensuelles la valeur d'une variable en  $t$  dépende de sa valeur en  $t - 12$  via :  $y_t = \psi y_{t-12} + u_t$ . Il s'agit d'un processus saisonnier simple, avec un seul coefficient autorégressif non nul, que l'on peut aussi présenter comme un AR(12) troué.

6. Pour reprendre l'exemple de la note précédente, la prise en compte de la saisonnalité ne doit pas se faire au moyen d'une modélisation de type AR dont l'ordre serait poussé au moins à 12, au risque de pénaliser fortement les filtres prenant en compte cette saisonnalité mais via des processus SARMAxARMA spécifiques, par exemple ici grâce à un processus n'ayant qu'un seul coefficient autorégressif à estimer.

7. Ici trop doit être en particulier compris en fonction de la longueur de la série traitée. En pratique, sur des échantillons de taille typique des séries macroéconomiques, au-delà de 2 coefficients MA on rencontre régulièrement ce type de difficultés.

## 1.2 L'intuition pour la construction d'un critère

On se trouve donc à l'issue de l'étape précédente amené à choisir entre un ensemble de modèles concurrents et pour cela nous disposons d'une seule trajectoire de la variable considérée, trajectoire observée sur  $T$  observations.

Dans ces conditions, et très logiquement, on conçoit que la vraisemblance associée à chaque filtre concurrent mérite d'être prise en compte : a priori, entre deux filtres, il semble raisonnable de retenir comme processus générateur potentiel celui qui est le plus à même d'avoir généré la trajectoire observée, ou en d'autres termes, celui pour lequel cette trajectoire est la plus vraisemblable.

Pour autant la seule considération de la vraisemblance conduit à des choix peu satisfaisants. On peut s'en apercevoir en considérant le tableau précédent : Tous les  $(p_{max} + 1)(q_{max} + 1)$  filtres sont emboîtés dans le plus grand processus considéré qui est le ARMA( $p_{max}, q_{max}$ ). En effet n'importe lequel des autres processus peut être obtenu en imposant des contraintes de nullité sur les coefficients des polynômes autorégressif ou moyenne mobile de cet ARMA( $p_{max}, q_{max}$ ). En conséquence le pouvoir explicatif du plus grand processus considéré est obligatoirement supérieur ou égal à ceux des processus qui s'en déduisent par imposition de contraintes sur les coefficients<sup>8</sup>. Ceci équivaut à dire que la vraisemblance du processus englobant constitue une borne supérieure pour l'ensemble des valeurs des vraisemblances des autres modèles. Dans ces conditions, un critère fondé sur la seule vraisemblance retiendrait certainement le modèle ARMA( $p_{max}, q_{max}$ ). Pour le coup la sélection des ordres  $p_{max}$  et  $q_{max}$ , qui initialement visait simplement à définir l'espace de recherche des ordres optimaux définirait en fait ces ordres optimaux, ce qui n'est évidemment pas l'objectif recherché<sup>9</sup>. Il importe donc de prendre en compte un autre aspect que la seule vraisemblance pour sélectionner un modèle optimal.

Supposons maintenant que deux modèles  $M_1$  et  $M_2$  ajustés sur un échantillon commun aient la même valeur pour leur fonction de vraisemblance, mais que  $M_1$  nécessite l'estimation d'un nombre  $k_1$  de paramètres inférieur au nombre de paramètres à estimer sous  $M_2$ , soit  $k_2 > k_1$ . De ce point de vue,  $M_1$  est donc un modèle plus parcimonieux que  $M_2$ . La question est de savoir quelle explication doit être privilégiée si l'on doit opérer un choix entre les deux représentations ? La réponse généralement acceptée est qu'il est préférable de prendre l'explication la plus simple, or dans le cas des processus ARMA, on peut imaginer que la complexité soit résumée par le nombre de paramètres. Ainsi un AR(5) est plus complexe qu'un AR(2) ou qu'un MA(3) ou qu'un ARMA(1,1). En revanche, toujours de ce point de vue, un AR(2) est aussi complexe qu'un MA(2) ou qu'un ARMA(1,1). Cette acception peut surprendre, ainsi un MA(2) inversible correspond à un AR( $\infty$ ) et il n'est donc pas si trivial que le MA(2) soit considéré comme moins complexe qu'un AR(4) par exemple. L'important est ici le nombre de paramètres à estimer qui est supposé résumer entièrement le degré de complexité du modèle<sup>10</sup> et naturellement cette position sera acceptée dans la suite de l'exposé.

Ainsi un critère de sélection sera fondé d'une part sur une mesure de vraisemblance et d'autre part sur une mesure de pénalité, cette dernière étant une fonction croissante avec le nombre de paramètres à estimer. Au total, si on note  $l_i()$  la log-vraisemblance du  $i^{ime}$  modèle et  $k_i$  son nombre de paramètres et  $T$  la taille de l'échantillon de travail, la majorité des critères de sélection va avoir comme écriture :

$$c_i(T, k_i) = \frac{-2l_i}{T} + \frac{k_i g(T)}{T} \quad (1)$$

Dans cette écriture, la fonction à valeurs positives  $g(T)$  gouverne la façon dont est pris en compte le nombre de paramètres à estimer et donc finalement détermine l'ampleur de la pénalité. Comme la vraisemblance est considérée négativement et que la pénalité est positivement liée à  $k$ , la règle de décision sera donc de choisir le modèle associé à la valeur du critère la plus faible<sup>11</sup>, soit si on a  $m$  modèles concurrents :

$$\hat{i} = \operatorname{argmin}_{i \leq m} c_i(T, k_i) \quad (2)$$

Fondamentalement le principe est qu'en augmentant la complexité on augmente la vraisemblance. Toute la question est de savoir si cette augmentation de la complexité est compensée par une augmentation suffisante de la vraisemblance. Si la réponse est négative alors on retient le modèle le moins complexe, même s'il est moins vraisemblable. Si la réponse est affirmative alors on accepte de travailler avec un modèle plus complexe. Bien évidemment toute la difficulté réside

8. Pour mémoire vous avez déjà rencontré ce type de résultat en étudiant les OLS : la qualité d'un ajustement mesurée par le  $R^2$  augmente même si on ajoute à un modèle des variables explicatives non pertinentes.

9. Clairement, choisir par exemple  $p_{max} = 5$  et  $q_{max} = 3$  ne signifie pas que l'on considère le filtre ARMA(5,3) comme étant optimal, mais simplement que l'on pense a priori que les ordres AR et MA optimaux sont inférieurs respectivement à 5 et 3.

10. Toujours afin d'illustrer l'éventuelle difficulté ainsi posée, imaginez une équation non linéaire à deux paramètres du type  $y_t = ax_t^b + u_t$ . Est-on certain qu'elle est moins complexe que l'équation à trois paramètres  $y_t = \beta_1 + \beta_2 x_t + \beta_3 x_{t-1} + u_t$  ?

11. par exemple, avec deux modèles tels que  $l_1 = l_2$  et  $k_1 < k_2$  on a  $c_1 < c_2$ , ce qui conduit bien à retenir le plus parcimonieux entre deux modèles ayant la même vraisemblance.

---

Akaike :	$c_i(T, k_i) = \frac{-2l_i}{T} + \frac{2k_i}{T}$
Schwarz :	$c_i(T, k_i) = \frac{-2l_i}{T} + \frac{k_i \log(T)}{T}$
Hannan-Quinn :	$c_i(T, k_i) = \frac{-2l_i}{T} + \frac{2k_i \log(\log(T))}{T}$

---

TABLE 2 – Les expressions de définition des trois critères

dans le choix des pondérations entre vraisemblance et complexité, et donc en définitive sur la spécification de la fonction de pénalité  $k_i g(T)$ .

### 1.3 Les critères de sélection les plus courants : Akaike, Schwarz, Hannan et Quinn

En pratique la quasi-totalité des travaux utilisant un critère de sélection font appel soit au critère d’Akaike (1974, 1974) qui est le plus populaire, au critère de Schwarz (1978) ou, dans une moindre mesure, au critère proposé par Hannan et Quinn (1979).

Leurs expressions sont présentées dans la table 2.

Avec le critère AIC d’Akaike<sup>12</sup> on a  $g(T) = 2$ , le critère  $\phi$  ou BIC de Schwarz<sup>13</sup> se caractérise par  $g(T) = \log(T)$ , et enfin le critère d’Hannan-Quinn retient<sup>14</sup>  $g(T) = 2\log(\log(T))$ . On se souviendra en particulier qu’avec n’importe lequel de ces critères on vérifie  $\lim_{T \rightarrow \infty} g(T)/T = 0$ .

Dans la mesure où la fonction de pénalité diffère selon les critères, vous devez être conscients que l’emploi de l’un ou de l’autre peut conduire à sélectionner des modèles différents. En d’autres termes **il n’existe pas dans cette logique de modèle optimal en termes absolus : un modèle est optimal seulement relativement au critère que l’on a utilisé pour le sélectionner.**

Compte-tenu de la remarque précédente, le point qu’il convient alors d’approfondir est celui du choix du critère : existe-t-il un critère qui soit lui-même optimal? Sans surprise la réponse est négative. On peut cependant concevoir des configurations dans lesquelles un critère se comporte a priori mieux, ou moins bien, que les autres. Pour en discuter, il convient de préciser quelque peu la consistance des choix auxquels ils aboutissent. On se situera dans un cadre simplifié où très logiquement la complexité d’un modèle est calée sur le nombre de paramètres dont il demande l’estimation, mais nous supposons aussi que le modèle à  $k_i$  paramètres est strictement emboîté dans le modèle à  $k_i + 1$  paramètres. C’est par exemple le cas si on limite notre recherche de modèle optimal à la seule famille des processus AR : l’AR(1) est en effet emboîté dans l’AR(2) et se déduit de celui-ci par l’imposition d’une contrainte de nullité sur l’un des paramètres autorégressifs ; de même l’AR(2) est emboîté dans l’AR(3), etc. Ce n’est évidemment pas le cas dans la famille plus large des ARMA : le processus ARMA(1,1), à deux paramètres n’est pas emboîté dans le processus AR(3). Par ailleurs dans ce cadre plus large, un même nombre de paramètres peut renvoyer à des modèles extrêmement différents<sup>15</sup>. Dans notre système simplifié, à un nombre de paramètre donné correspond un modèle et un seul<sup>16</sup>. Dans ces conditions, chaque modèle se distingue par son nombre de paramètres. Nous allons également supposer que le vrai modèle est dans la famille des modèles concurrents : si  $k$  est la taille du modèle le plus complexe fixé a priori alors nous avons  $k$  modèles concurrents, chacun ayant un nombre de paramètres  $i$ , avec  $i = 0, 1, \dots, k$ , et parmi ceux-ci le  $j^{\text{ième}}$  est le vrai avec  $j < k$ . Toute la question

12. Les lettres IC de AIC renvoient à *Information Criteria*. Le point de départ de l’analyse d’Akaike est la distance de Kullback-Leibler entre deux fonctions de densité  $I(f; g) = \int f(x) \log \frac{f(x)}{g(x)} dx$ , où  $f$  est la vraie densité et  $g$  la densité supposée. Akaike montre que AIC est une approximation de cette distance.

13. le B traduit le fait non pas qu’il ait été le second critère d’information proposé après celui Akaike, mais que sa dérivation est réalisée dans un cadre d’analyse bayésien.

14. En fait Hannan et Quinn proposent  $c_i(T, k_i, c) = \frac{-2l_i}{T} + \frac{c k_i \log(\log(T))}{T}$ , où  $c$  est une constante devant vérifier  $c > 1$ . On retient ici la valeur la plus couramment imposée,  $c = 2$ .

15. considérez par exemple les processus ARMA(2,2), ARMA(3,1), ARMA(1,3), AR(4), MA(4), .... Au passage, remarquez que les critères vont pouvoir opérer une sélection entre ces filtres mais que celle-ci sera fondée sur la seule comparaison des vraisemblances puisqu’ils ont des mesures de pénalité identiques.

16. Si vous reprenez la seule famille des AR, et hors prise en compte d’un éventuel terme constant, un modèle à un paramètre est un AR(1), si on a deux paramètres alors on est en présence d’un AR(2) qui englobe le précédent, etc...

est donc de savoir si le vrai modèle est aussi optimal ou, en d'autres termes, si les critères vont sélectionner le vrai modèle ?

Afin de traiter ce problème nous allons distinguer deux cas : quelle est la probabilité (asymptotique) de sélectionner un modèle de taille inférieure au vrai modèle ? Puis, quelle est la probabilité (asymptotique) de sélectionner un modèle de taille supérieure au vrai modèle ? Ceci revient donc à évaluer, pour un critère donné, la consistance de l'estimateur optimal de l'ordre vrai inconnu.

### 1.3.1 probabilité de sélection d'un ordre inférieur à l'ordre vrai

Si  $p^*$  est l'ordre vrai, alors puisque le modèle d'ordre  $p$  comporte omet des explicatives pertinentes, nous avons :

$$plim_{T \rightarrow \infty} l_p < plim_{T \rightarrow \infty} l_{p^*},$$

soit encore :

$$plim_{T \rightarrow \infty} \frac{l_p}{T} < plim_{T \rightarrow \infty} \frac{l_{p^*}}{T} \quad (3)$$

Au moyen de (1) et (3) nous pouvons évaluer la probabilité asymptotique de sélection d'un modèle dont l'ordre est inférieur à l'ordre vrai, qui est la probabilité que la valeur du critère afférent au modèle  $p$  soit inférieure à la valeur du critère obtenu sur le modèle  $p^*$  :

$$\begin{aligned} \lim_{T \rightarrow \infty} P [c(T, p) \leq c(T, p^*)] &= \\ &= \lim_{T \rightarrow \infty} P \left[ \frac{-2l_p}{T} + \frac{p g(T)}{T} \leq \frac{-2l_{p^*}}{T} + \frac{p^* g(T)}{T} \right] \\ &= \lim_{T \rightarrow \infty} P \left[ \frac{l_{p^*}}{T} - \frac{l_p}{T} \leq \frac{(p^* - p) g(T)}{2T} \right] \\ &= 0, \end{aligned} \quad (4)$$

la dernière égalité étant impliquée par (3) et le fait que pour chacun des 3 critères considérés on ait  $\lim_{T \rightarrow \infty} g(T)/T = 0$ .

On peut donc retenir qu'avec les critères AIC, BIC ou HQ, la probabilité asymptotique de sous-estimer l'ordre vrai est nulle.

### 1.3.2 probabilité de sélection d'un ordre supérieur à l'ordre vrai

Dans cette configuration le modèle sélectionné est surparamétré : des paramètres, en nombre  $p - p^*$ , sont estimés alors que leur vraie valeur est nulle. Comme on sait estimer les deux modèles à  $p$  et  $p^*$  paramètres, il est possible de considérer le test LRT de nullité des paramètres en trop, l'avantage étant qu'alors on connaît la distribution de l'écart des vraisemblances. Plus précisément :

$$LRT_{p,p^*} = 2(l_p - l_{p^*}) \rightarrow_d \chi^2(p - p^*) \quad (5)$$

En reprenant l'équation de définition générale des critères de sélection (1), il vient immédiatement :

$$T [c(T, p^*) - c(T, p)] = 2(l_p - l_{p^*}) + (p^* - p) g(T) = z + (p^* - p) g(T),$$

où  $z$  est une variable de Chi2 à  $(p - p^*)$  degrés de liberté, et donc :

$$\lim_{T \rightarrow \infty} P [c(T, p^*) > c(T, p)] = \lim_{T \rightarrow \infty} P [z > (p - p^*) g(T)] \quad (6)$$

A ce stade il va être nécessaire de distinguer la signification de (6) selon les définitions retenues pour  $g(T)$ .

— avec le critère AIC d'Akaike :  $g(T) = 2$ . En conséquence la probabilité asymptotique de sélectionner le modèle surparamétré n'est pas nulle puisqu'elle est égale à la probabilité qu'une variable de Chi2 à  $(p - p^*)$  degrés de liberté soit supérieure à une valeur positive finie égale à  $2(p - p^*)$ .

— en revanche, avec  $g(T) = \log(T)$  pour le critère de Schwarz, et  $g(T) = 2\log(\log(T))$  pour le critère de Hannan et Quinn, on vérifie dans les deux cas que  $\lim_{T \rightarrow \infty} g(T) = +\infty$ . En conséquence,

$$\begin{aligned} plim \frac{T [c(T, p^*) - c(T, p)]}{g(T)} &= plim \frac{z}{g(T)} + (p^* - p) \\ &= p^* - p < 0 \end{aligned}$$

et donc,

$$\lim_{T \rightarrow \infty} P [c(T, p^*) \geq c(T, p)] = 0$$

On a donc avec ces deux critères une probabilité asymptotique nulle de sélectionner un modèle surparamétré.

Pour résumer, la probabilité asymptotique de sélectionner un modèle d'ordre inférieur à l'ordre vrai est nulle pour les trois critères<sup>17</sup>. En revanche la probabilité de sélectionner un modèle d'ordre supérieur à l'ordre vrai n'est pas nulle avec le critère d'Akaike alors qu'elle est nulle pour les deux autres critères<sup>18</sup>. On peut en conclure naturellement que la probabilité asymptotique de sélectionner le modèle vrai est égale à l'unité lorsqu'on emploie les critères de Schwarz ou de Hannan et Quinn.

Étant donné le précédent résultat on peut se demander pour quelles raisons le critère d'Akaike continue d'être le plus populaire? Tout d'abord il faut toujours se rappeler que ce résultat est seulement asymptotique : il n'est pas certain qu'il soit très utile à distance finie. Cet aspect a été étudié dans la littérature au moyen notamment de simulations de Monte Carlo. Nous en donnerons quelques exemples dans la section suivante, les résultats étant représentatifs de ceux obtenus dans la majorité des études.

Par ailleurs nous avons supposé dans les précédents développements que le modèle vrai était présent au sein de la famille de modèles concurrents. Il est tout à fait possible que cette condition ne soit pas vérifiée. Cette situation arrive de manière triviale lorsque l'ordre maximal choisi a priori est inférieur à l'ordre vrai, ce qui revient à dire que le modèle vrai ne peut être retrouvé en imposant des contraintes sur les paramètres du modèle le plus complexe<sup>19</sup>. Ce cas n'est peut être pas trop inquiétant puisque l'on peut espérer que le critère va sélectionner ce modèle le plus complexe et donc ainsi signaler la nécessaire augmentation de l'espace de recherche des ordres optimaux<sup>20</sup>. La non présence du modèle vrai parmi les modèles concurrents peut cependant arriver dans un autre cas sans que l'augmentation des ordres de recherche puisse corriger la difficulté. Il est même possible que l'utilisateur soit conscient de cette absence. Pour illustrer cette configuration, supposez que pour une raison quelconque on limite la recherche du modèle optimal à la classe des processus AR(p) et que l'on sache que le processus vrai est un AR( $\infty$ ). A l'évidence, et quelle que soit la valeur fixée pour  $p_{max}$ , le modèle vrai ne peut appartenir à la famille  $\{AR(0), AR(1), AR(2), \dots, AR(p_{max})\}$ . La question n'est donc pas de savoir si le modèle optimal est le modèle vrai, mais plutôt de rechercher la meilleure approximation par un AR d'ordre fini du processus AR( $\infty$ ). Dans ces conditions la propriété de consistance associée aux critères de Schwarz ou d'Hannan et Quinn n'est plus d'aucune utilité. Bien plus, l'emploi d'un critère dont on sait qu'il retient en moyenne plus de retards que les deux précédents peut apparaître comme préférable dans cette recherche d'une approximation finie satisfaisante. Ce raisonnement a été formalisé notamment par Shibata<sup>21</sup> qui introduit le concept d'efficacité asymptotique : il s'agit de sélectionner un AR fini, approximation d'un AR( $\infty$ ) qui minimise l'erreur quadratique moyenne. Il montre que pour des résidus gaussiens le critère AIC possède cette propriété d'efficacité asymptotique contrairement à BIC et HQ<sup>22</sup>. Il n'est donc pas surprenant que dans cette situation le critère d'Akaike soit alors privilégié<sup>23</sup>

17. notamment en raison de la propriété  $\lim_{T \rightarrow \infty} g(T)/T = 0$ .

18. notamment en raison de la propriété  $\lim_{T \rightarrow \infty} g(T) = +\infty$ .

19. Supposez ainsi que le vrai processus soit un AR(8) mais que vous ayez fixé  $p_{max} = 5$ .

20. Dans l'exemple précédent, si le critère sélectionne le processus AR(5), et donc l'ordre optimal est égal à  $p_{max}$ , cela impose évidemment de refaire une recherche avec une nouvelle valeur de  $p_{max}$  supérieure à 5.

21. Shibata, R. (1980). Asymptotically efficient selection of the order of the model for estimating parameters of a linear process. *Ann. Statist.* **8**, 147-164.

22. Ce résultat est étendu aux processus à résidus non gaussiens par Karagrigoriou (in Karagrigoriou, A. (1997). Asymptotic efficiency of the order selection of a nongaussian AR process. *Statistica Sinica* **7**, 407-423.

23. Il est vrai que l'emploi d'un autre concept que l'efficacité asymptotique peut ne pas confirmer cette conclusion. Ainsi Galbraith et Zinde-Walsh utilisent la distance de Hilbert pour définir un processus AR(p) fini optimal lorsque le vrai processus est un AR( $\infty$ ). Cette distance entre 2 processus X et Y stationnaires est donnée par  $d_H(X, Y) = \|XY\| = [E(X - Y)^2]^{1/2}$ . Si par exemple le vrai filtre X est un AR( $\infty$ ), et il suffit pour cela qu'il soit un ARMA(p,q>0) inversible, alors il est possible de trouver dans la classe des AR d'ordre fini le processus Y d'ordre  $p^*$  qui minimise cette distance. Ce processus Y est qualifié d'optimal. La question suivante est de juger des différents critères de sélection à partir de leur capacité à retrouver ce modèle optimal. Les auteurs, au moyen de cas simulés, concluent à la supériorité du critère BIC et, à un moindre degré, de HQ sur AIC (in Galbraith, J. W. and Zinde-Walsh, V. (2004). Évaluation de critères d'information pour les modèles de séries chronologiques. *Revue d'Analyse Économique* Volume **80**, numéros 2-3, 207-227).

## 1.4 Illustrations

Dans cette partie nous donnons simplement les résultats de simulations obtenus pour des configurations très simples, l'objectif étant seulement d'illustrer divers aspects abordés précédemment.

Les diverses configurations étudiées sont présentées ci-après. On indique le processus utilisé pour générer la trajectoire, ou DGP (*Data Generating Process*) : il s'agit du vrai modèle. Les ordres de recherche considérés sont définis par les valeurs des paramètres  $p_{max}$  et  $q_{max}$ .

1. DGP : AR(1) selon  $y_t = .8y_{t-1} + u_t$  ;  $p_{max} = 4$ ,  $q_{max} = 0$
2. DGP : AR(2) selon  $y_t = .4y_{t-1} + .4y_{t-2} + u_t$  ;  $p_{max} = 4$ ,  $q_{max} = 0$
3. DGP : MA(2) selon  $y_t = u_t + .4u_{t-1} + .4u_{t-2}$  ;  $p_{max} = 0$ ,  $q_{max} = 4$
4. DGP : ARMA(1,1) selon  $y_t = .8y_{t-1} + u_t + .8u_{t-1}$  ;  $p_{max} = 4$ ,  $q_{max} = 4$

Dans la totalité des expériences la variable  $u_t$  est la réalisation pseudo-aléatoire d'une gaussienne centrée réduite. Les valeurs initiales pour  $y_1$  dans les cas 1 et 5, et pour  $y_1$  et  $y_2$  dans le cas 2 sont égales à zéro<sup>24</sup>. Chaque DGP est simulé 10000 fois sur une longueur  $T$ <sup>25</sup>. Différentes tailles d'échantillon sont considérées pour repérer d'éventuelles divergences entre les propriétés attendues asymptotiquement et celles obtenues à distance finie ; ainsi  $T \in \{30, 60, 180, 500\}$ .

Les tableaux suivants indiquent pour chaque modèle concurrent le pourcentage de fois où il a été sélectionné parmi les 10000 tirages. On donne également, lorsque cela est nécessaire, le pourcentage de fois où chacun de ces modèles a pu être estimé et donc où il n'a pas été rencontré de difficultés insurmontables de convergence dans la recherche des estimateurs des paramètres<sup>26</sup>.

1. Cas d'un DGP de type AR(1)

DGP : AR(1)		p				
		0	1	2	3	4
T=30	AIC	1%	<b>71%</b>	14%	7%	7%
	BIC	2%	<b>87%</b>	7%	2%	1%
	HQ	1%	<b>79%</b>	11%	5%	4%
T=60	AIC	0%	<b>75%</b>	13%	7%	6%
	BIC	0%	<b>94%</b>	5%	1%	0%
	HQ	0%	<b>86%</b>	9%	3%	2%
T=180	AIC	0%	<b>75%</b>	13%	7%	6%
	BIC	0%	<b>97%</b>	2%	0%	0%
	HQ	0%	<b>90%</b>	7%	2%	1%
T=500	AIC	0%	<b>77%</b>	12%	6%	5%
	BIC	0%	<b>98%</b>	2%	0%	0%
	HQ	0%	<b>93%</b>	5%	1%	0%

On peut noter que le nombre de processus sélectionnés dont l'ordre est supérieur à l'unité a tendance à diminuer pour tous les critères lorsque le nombre d'observations augmente. Cependant, même avec 500 observations, le critère AIC continue de retenir des processus sur-dimensionnés dans 23% des cas alors que BIC et HQ désigne le bon processus dans respectivement 98% et 93% des expériences. Ceci est naturellement conforme au résultat montrant que la probabilité de surestimer l'ordre vrai ne s'annule pas avec l'AIC alors qu'elle tend vers zéro avec les deux autres critères.

24. Ce qui revient à leur donner la valeur de l'espérance non conditionnelle de  $y_t$ .

25. En fait la série est simulée sur  $T+T_0$  observations, mais seules les  $T$  dernières sont prises en compte. Ceci a pour objet de limiter l'influence des conditions initiales arbitrairement imposées (Cf. la note précédente) sur les réalisations considérées dans les calculs. Ici nous avons utilisé  $T_0 = 50$

26. Ici le terme *insurmontable* renvoie à l'impossibilité pour le logiciel de satisfaire un critère de convergence. Un constat de non convergence peut donc être associé naturellement à des choix malheureux de valeurs initiales lorsque des procédures itératives sont mises en oeuvre, et aux algorithmes de maximisation utilisés. Dans notre cas, les calculs ont été effectués sous RATS 7 avec des conditions initiales pour les paramètres qui sont celles utilisées par défaut par ce logiciel. D'après notre expérience, les routines de maximisation implémentées dans RATS se comparent sans problème avec celles d'un logiciel tel que Matlab, avec l'avantage d'être souvent plus rapides.

## 2. Cas d'un DGP de type AR(2)

Cet exercice permet de retrouver les caractéristiques soulignées dans le cas précédent. Par ailleurs il permet de s'intéresser à la probabilité de sélection d'un processus sous-estimant les ordres vrais. On sait qu'asymptotiquement cette probabilité est nulle pour les trois critères et c'est d'ailleurs bien ce qui apparaît sur les deux derniers échantillons. En revanche, sur petits échantillons, la sous-estimation n'est pas un événement rare. Par exemple le processus optimal est un AR(1) avec le critère BIC dans près de 20% des cas pour des séries de 60 observations et dans près de 40% si on ne dispose que de 30 observations. Sur ces tailles d'échantillons c'est d'ailleurs le critère AIC qui conduit le moins souvent à une sous-estimation de l'ordre vrai. Sachant que souvent les conséquences sur les propriétés des estimateurs sont plus nocives lorsque l'on omet des variables explicatives pertinentes que lorsque l'on retient des explicatives non pertinentes, on peut imaginer que ce type de résultats conduise à préférer l'emploi de AIC.

DGP : AR(2)		p				
		0	1	2	3	4
T=30	AIC	11%	28%	<b>42%</b>	11%	8%
	BIC	21%	<b>39%</b>	34%	4%	2%
	HQ	14%	33%	<b>40%</b>	8%	5%
T=60	AIC	1%	8%	<b>70%</b>	12%	9%
	BIC	3%	22%	<b>71%</b>	3%	1%
	HQ	1%	14%	<b>74%</b>	7%	4%
T=180	AIC	0%	0%	<b>78%</b>	14%	8%
	BIC	0%	0%	<b>97%</b>	2%	0%
	HQ	0%	0%	<b>91%</b>	7%	2%
T=500	AIC	0%	0%	<b>78%</b>	13%	9%
	BIC	0%	0%	<b>99%</b>	1%	0%
	HQ	0%	0%	<b>93%</b>	5%	2%

## 3. Cas d'un DGP de type MA(2)

Pour l'essentiel il s'agit ici de confirmer les résultats observés au point précédent. La différence majeure pourrait provenir du fait que les processus MA sont plus difficile à estimer que les processus AR<sup>27</sup>. Nous avons ainsi fait apparaître une ligne ("n") précisant le nombre de processus pour lesquels, parmi les 10000 simulés, l'étape d'estimation a été menée à son terme avec succès.

Comme on pouvait l'attendre, c'est surtout sur les faibles tailles d'échantillon que des difficultés de convergence sont rencontrées, leur nombre étant, également très logiquement, croissant avec le nombre de paramètres à estimer. Pour le reste, on retrouve effectivement des résultats conformes à ceux de l'expérience précédente : mise en évidence de la consistance asymptotique avec les critères BIC et HQ, probabilité de sous-estimation de l'ordre vrai qui tend vers zéro pour les trois critères, et probabilité asymptotique non nulle de surestimation avec AIC. On note également à nouveau que le risque de sous-estimation, présent pour des échantillons de taille réduite, est toujours le plus faible avec AIC.

<sup>27</sup>. On rappelle que les estimateurs du maximum de vraisemblance conditionnels des AR sont simplement les estimateurs OLS. En revanche les processus ARMA ou MA imposent le recours à des algorithmes itératifs de maximisation et peuvent donc connaître des difficultés de convergence

DGP : MA(2)		q				
		0	1	2	3	4
T=30	AIC	17%	15%	<b>42%</b>	14%	12%
	BIC	36%	18%	<b>35%</b>	6%	5%
	HQ	23%	16%	<b>41%</b>	11%	9%
	n	10000	9977	9902	9245	8076
T=60	AIC	2%	5%	<b>67%</b>	15%	11%
	BIC	12%	13%	<b>69%</b>	4%	2%
	HQ	5%	8%	<b>72%</b>	9%	6%
	n	10000	10000	9995	9961	9802
T=180	AIC	0%	0%	<b>77%</b>	13%	9%
	BIC	0%	0%	<b>97%</b>	3%	0%
	HQ	0%	0%	<b>90%</b>	7%	3%
	n	10000	10000	10000	10000	10000
T=500	AIC	0%	0%	<b>78%</b>	14%	8%
	BIC	0%	0%	<b>99%</b>	1%	0%
	HQ	0%	0%	<b>93%</b>	5%	2%
	n	10000	10000	10000	10000	10000

#### 4. Cas d'un DGP de type ARMA(1,1)

Il s'agit donc maintenant de retrouver un processus vrai dans une famille plus complexe que celles précédemment considérées puisque maintenant on considère simultanément une dimension autorégressive et une dimension moyenne mobile.

Comme on peut le voir d'après le nombre de modèles qui ont pu être ajustés, un processus ARMA( $p > 0, q > 0$ ) est effectivement plus difficile à estimer que les processus de type AR ou MA <sup>28</sup>.

On remarque encore que la probabilité de sous-estimer les ordres vrais tend bien vers zéro avec AIC et BIC, et pour ces deux critères le mode de la distribution des ordres optimaux est bien obtenu pour le couple ( $p = 1, q = 1$ ) qui correspond au vrai processus. La consistance de l'estimation BIC étant d'ailleurs assez remarquable : déjà avec 60 points il sélectionne le bon filtre dans 73% des cas (44% pour AIC), ce pourcentage montant à 92% avec des séries de longueur 160 (46% pour AIC). En revanche dans ces essais, le critère HQ ne parvient pas à faire entrer de composante MA dans le modèle optimal.

La leçon à retenir est assez simple : il faut un nombre d'observations important pour se permettre d'ajuster des processus ARMA(p,q) même d'ordres relativement faibles. Si on pense être en droit de faire ces ajustements et si le modèle vrai est dans la famille des modèles concurrents le critère BIC n'est sans doute pas un mauvais choix. Un autre raisonnement, déjà énoncé à la fin de la section précédente, peut aussi être le suivant : puisque les ARMA(p,q) sont difficiles à estimer, considérons leur écriture équivalente AR( $\infty$ ), et recherchons dans cette classe de processus, qui ne pose pas de problèmes majeurs d'estimation, la meilleure approximation AR( $p$ ), avec  $p < \infty$  <sup>29</sup>. On peut donc alors comprendre la popularité des processus AR d'une part et l'emploi majoritaire du critère AIC d'autre part.

28. Le cas limite de ces expériences est caricatural puisque la convergence est obtenue avec seulement 99 ajustements de ARMA(4,4) sur les 10000 essayés pour des séries constituées de 30 observations. Mais, même avec 500 observations, l'ajustement de processus ARMA(4,4) échoue encore dans près de 7000 cas.

29. En faisant alors référence au concept d'efficacité asymptotique défini par Shibata présenté plus haut.

DGP : ARMA(1,1)

T=30		q						HQ		q					
AIC		0	1	2	3	4			0	1	2	3	4		
p	0	0%	1%	4%	2%	9%	p	0	16%	0%	0%	0%	0%		
	1	4%	<b>28%</b>	5%	2%	4%		1	<b>41%</b>	0%	0%	0%	0%		
	2	9%	6%	1%	2%	1%		2	21%	0%	0%	0%	0%		
	3	8%	1%	1%	0%	0%		3	10%	0%	0%	0%	0%		
	4	13%	3%	1%	0%	0%		4	13%	0%	0%	0%	0%		
BIC		q						n		q					
		0	1	2	3	4			0	1	2	3	4		
p	0	0%	1%	6%	2%	6%	p	0	10000	7676	7514	5159	5050		
	1	2%	<b>39%</b>	3%	1%	2%		1	10000	8066	7247	4955	3755		
	2	17%	4%	1%	1%	1%		2	10000	6726	2411	1937	993		
	3	6%	1%	0%	0%	0%		3	10000	6089	2432	608	248		
	4	6%	0%	0%	0%	0%		4	10000	5489	1790	452	99		

T=60		q						HQ		q					
AIC		0	1	2	3	4			0	1	2	3	4		
p	0	0%	0%	0%	1%	5%	p	0	6%	0%	0%	0%	0%		
	1	0%	<b>44%</b>	5%	2%	3%		1	<b>66%</b>	0%	0%	0%	0%		
	2	1%	7%	4%	3%	3%		2	13%	0%	0%	0%	0%		
	3	3%	2%	1%	2%	1%		3	7%	0%	0%	0%	0%		
	4	7%	3%	1%	1%	0%		4	8%	0%	0%	0%	0%		
BIC		q						n		q					
		0	1	2	3	4			0	1	2	3	4		
p	0	0%	0%	1%	1%	3%	p	0	10000	8599	9271	7843	8215		
	1	0%	<b>73%</b>	3%	0%	1%		1	10000	9444	9228	8261	7667		
	2	3%	4%	1%	1%	1%		2	10000	8713	4472	4697	3861		
	3	3%	0%	0%	0%	0%		3	10000	8432	5193	2165	1520		
	4	3%	0%	0%	0%	0%		4	10000	7867	4381	1902	698		

DGP : ARMA(1,1)

T=180		q						HQ		q					
AIC		0	1	2	3	4			0	1	2	3	4		
p	0	0%	0%	0%	0%	0%	p	0	0%	0%	0%	0%	0%		
	1	0%	<b>46%</b>	4%	2%	2%		1	<b>79%</b>	0%	0%	0%	0%	0%	
	2	0%	5%	9%	2%	2%		2	12%	0%	0%	0%	0%	0%	
	3	0%	2%	2%	11%	3%		3	7%	0%	0%	0%	0%	0%	
	4	0%	2%	1%	3%	3%		4	3%	0%	0%	0%	0%	0%	
BIC		q						n		q					
		0	1	2	3	4			0	1	2	3	4		
p	0	0%	0%	0%	0%	0%	p	0	10000	9711	9958	9663	9911		
	1	0%	<b>92%</b>	1%	0%	0%		1	10000	9996	9989	9948	9935		
	2	0%	2%	2%	0%	0%		2	10000	9908	7304	7870	7451		
	3	0%	0%	0%	1%	0%		3	10000	9913	8500	6059	5577		
	4	0%	0%	0%	0%	0%		4	10000	9796	7739	5674	3250		

T=500		q						HQ		q					
AIC		0	1	2	3	4			0	1	2	3	4		
p	0	0%	0%	0%	0%	0%	p	0	0%	0%	0%	0%	0%		
	1	0%	<b>41%</b>	3%	2%	1%		1	<b>82%</b>	0%	0%	0%	0%	0%	
	2	0%	4%	9%	2%	1%		2	9%	0%	0%	0%	0%	0%	
	3	0%	2%	2%	17%	5%		3	7%	0%	0%	0%	0%	0%	
	4	0%	1%	1%	4%	6%		4	2%	0%	0%	0%	0%	0%	
BIC		q						n		q					
		0	1	2	3	4			0	1	2	3	4		
p	0	0%	0%	0%	0%	0%	p	0	10000	9991	9997	9954	9994		
	1	0%	<b>96%</b>	1%	0%	0%		1	10000	10000	10000	10000	10000		
	2	0%	1%	1%	0%	0%		2	10000	9998	8650	9002	8717		
	3	0%	0%	0%	0%	0%		3	10000	9997	9380	8143	7778		
	4	0%	0%	0%	0%	0%		4	10000	9994	8983	7967	5808		

Minimum Information Criterion				
Lags	MA 0	MA 1	MA 2	MA 3
AR 0	5.280312	5.038683	4.818006	4.548886
AR 1	2.610385	1.865479	1.758647	1.731858
AR 2	1.515689	1.539418	1.585086	1.626039
AR 3	1.529742	1.574472	1.616015	1.661374

Error series model: AR(9)

Minimum Table Value: BIC(2,0) = 1.515689

TABLE 3 – Exemple de sortie produite par l’option `minic`

## 2 Leur mise en oeuvre sous SAS et R

### 2.1 L’option `minic` de SAS

#### 2.1.1 la démarche mise en oeuvre par `minic`

Dans la commande `identify` de la `proc arima` on dispose de l’option `minic`. Elle permet de définir un ensemble de  $(p_{max} - p_{min} + 1) \times (q_{max} - q_{min} + 1)$  modèles ARMA(p,q),  $p = p_{min}, \dots, p_{max}$ ,  $q = q_{min}, \dots, q_{max}$ . Pour chacun elle va estimer la valeur du critère BIC de Schwarz. On sait que le modèle optimal est celui ayant le BIC minimum. L’ensemble des processus concurrents peut être précisé via les options `p=(pmin,pmax)` et `q=(qmin,qmax)`. Par défaut,  $p_{min} = q_{min} = 0$  et  $p_{max} = q_{max} = 5$ .

Notez que la procédure simplifie l’estimation de chaque modèle concurrent : sous hypothèse de composante MA inversible, on sait que tous ces modèles possèdent une écriture AR éventuellement infinie. Elle en recherche une approximation en ajustant un AR d’ordre élevé,  $p_{error}$ , dont elle récupère les résidus empiriques,  $\hat{\epsilon}_t$ . Ensuite, chacun des processus ARMA(p,q) est estimé en appliquant les OLS sur :

$$x_t = \sum_{i=1}^p \phi^{(p,i)} x_{t-i} + \sum_{j=1}^q \theta^{(q,j)} \hat{\epsilon}_{t-j} + u_t \quad (7)$$

pour  $p = p_{min}, \dots, p_{max}$ ,  $q = q_{min}, \dots, q_{max}$ .

La valeur du critère est alors calculée selon :

$$BIC(p, q) = \log\left(\frac{1}{T} \sum \hat{u}_t^2\right) + 2(p + q) \log(n)/n \quad (8)$$

La valeur de l’ordre du processus autorégressif initial,  $p_{error}$ , est choisie en utilisant le critère d’Akaike sur un ensemble de valeurs concurrentes  $p_{error_{min}}, \dots, p_{error_{max}}$  qui peuvent être spécifiées avec `perror=(perrormin : perrormax)`.

#### 2.1.2 un exemple

Sur une série X de 150 observations, l’exécution des lignes ci-après a produit la table 3. Le critère d’Akaike a donc tout d’abord sélectionné un AR(9) dans l’ensemble {AR(8),AR(9),...,AR(12)} dont le résidu a été mis en variable explicative dans les régressions décrites par l’équation (7) et finalement le critère de Schwarz considère que le processus AR(2) est optimal dans l’ensemble des 16 filtres ARMA définis par les choix faits avec `p=` et `q=`.

```
proc arima data=simul;
identify var=x minic p=(0:3) q=(0:3) perror=(8:12);
run;
quit;
```

## 2.2 Sous R : au moins deux commandes

### 2.2.1 la commande `armaselect` du package `Caschrono`

Le package `caschrono` écrit par Y. Aragon, accompagne son ouvrage "Séries Temporelles Avec R" (EDP Sciences - Collection : Pratique R, 2016). On y trouve notamment les séries et fonctions utilisées dans le livre. Parmi les commandes offertes, `armaselect` reproduit la démarche de l'option `minic` de SAS : c'est également le critère BIC de Schwarz, nommé ici SBC pour *Schwarz Bayesian Criteria*, qui est utilisé pour la sélection du modèle optimal. La seule différence concerne la restitution des résultats : ici seuls sont affichés les meilleurs modèles au sens du critère BIC. Naturellement il faudra toujours préciser la famille des modèles concurrents en précisant les valeurs maximales de `p` et `q` Si la variable `x` contient les observations que l'on veut modéliser, on fera par exemple :

```
armaselect(x,max.p=8,max.q=8,nbmod=3)
```

pour récupérer une liste présentant les paramètres `p`, `q` ainsi que la valeur du critère des 3 meilleurs processus :

	p	q	sbc
[1,]	6	1	16.76714
[2,]	5	1	20.06077
[3,]	7	1	21.01959

### 2.2.2 la commande `armasubsets` des packages `TSA` et `astsa`

Le package `TSA` a été construit pour accompagner l'ouvrage "Time Series Analysis with Applications in R - second edition" de Jonathan Cryer and Kung-Sik Chan (Springer Texts in Statistics). Le package `astsa` accompagne le manuel "Time Series Analysis and Its Applications With R Examples - 4th Edition" de Robert H. Shumway & David S. Stoffer (Springer, 2017). Je vous invite à consulter la page de l'un des auteurs (<https://www.stat.pitt.edu/stoffer/tsa4/>), elle offre des ressources qui pourraient vous intéresser.

Dans ces packages, la commande `armasubsets` utilise initialement la même démarche que `minic` de SAS : estimation d'un AR dont l'ordre est choisi par AIC, puis des régressions ols telles que définies par l'équation (7). Son grand intérêt relativement à `minic` et à `armaselect` de `Caschrono` est qu'elle va considérer des ARMA troués dans la famille des modèles concurrents. C'est donc tout particulièrement lorsque l'on effectue une recherche de modèle optimal dans la classe des processus SARMA qu'elle peut être préférée aux deux autres commandes.

Pour illustrer cet avantage, supposez que le modèle vrai soit un  $SAR(2)_{12} : x_t = \Phi_1 x_{t-12} + \Phi_2 x_{t-24} + u_t$ . C'est donc un AR(24) troué pour lequel on doit estimer seulement deux coefficients. Si, comme le font `minic` et `armaselect`, on effectue la recherche uniquement dans la classe des processus non troués, alors ce modèle vrai apparaît comme le cas particulier d'un AR(24), *i.e.* d'un processus à 24 coefficients. Sachant que la pénalité incorporée aux critères augmente avec le nombre de coefficients à estimer, il est fort possible que ces deux commandes ne sélectionnent pas le modèle duquel pourrait se déduire le vrai processus.

Attention, la commande `armasubsets` s'applique à une série temporelle : la variable a préalablement été créée par exemple par la commande `ts`, selon :

```
xser <- ts(x,start=1,end=240,frequency=12)
```

On peut alors l'utiliser, en précisant les valeurs maximales de `p` et `q` bornant la famille de modèles concurrents via les options `nar=` et `nma=` .

### 2.2.3 un exemple

Nous avons simulé un  $SMA(1) \times AR(1)_{12}$  d'écriture :  $x_t = 0.8x_{t-12} + u_t + 0.6u_{t-1}$ . La sélection d'un processus optimal est demandé pour des ordres `p` et `q` au plus égaux à 14 avec :

```
select <-armasubsets(y=x,nar=14,nma=14)
plot select
```

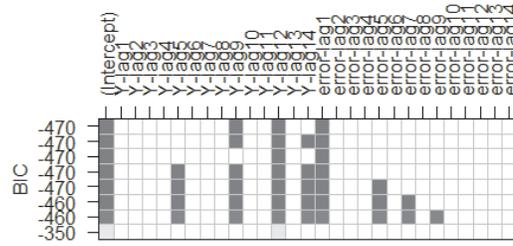


TABLE 4 – Table de résultats fourni par armasubsets

On récupère alors la table 4. Dans celle-ci, les modèles sont classés par ordre de BIC croissant : le meilleur modèle occupe la première ligne, le suivant la deuxième ligne, etc... Chaque colonne représente un paramètre : la première colonne correspond à la constante, ensuite viennent les coefficients des 14 ordres autorégressifs possibles, puis les coefficients afférents aux 14 retards de type MA. Un case noircie signifie que le coefficient est présent dans un modèle. Ainsi, le modèle optimal au sens du critère BIC comporte une constante,  $x_{t-9}$ ,  $x_{t-12}$  et un terme  $u_{t-1}$ . Au retard de 9 périodes sur  $x$  près, on retrouve donc sur cet exemple le modèle simulé. Remarquez également que dans les 8 meilleurs modèles retenus, on retrouve toujours les bonnes explicatives,  $x_{t-12}$  et  $u_{t-1}$ .