REVUE DE STATISTIQUE APPLIQUÉE

F. B. DOUCOURÉ

Sélection de méthodes par le critère de l'erreur quadratique moyenne de prédiction

Revue de statistique appliquée, tome 44, nº 3 (1996), p. 27-45
<http://www.numdam.org/item?id=RSA_1996__44_3_27_0>

Rev. Statistique Appliquée, 1996, **XLIV** (3), 27-45

SÉLECTION DE MÉTHODES PAR LE CRITÈRE DE L'ERREUR QUADRATIQUE MOYENNE DE PRÉDICTION

F.B. Doucouré

Université Paris XIII

Département de Mathématiques, Institut Galilée

Avenue J. B. Clément, 93430 Villetaneuse

RÉSUMÉ

Dans cette étude, nous utilisons le critère de l'erreur quadratique moyenne de prédiction pour comparer la performance de différentes méthodes de prédiction sur les modèles : ARMA, RCA, BILINEAIRE, ARCH, ARMA-ARCH et FARMA. Cette étude montre pourquoi certaines méthodes sont plus précises que d'autres sur de tels modèles. Nous mettons ainsi en évidence que pour certains modèles des méthodes intuitives comme la moyenne mobile simple sont meilleures que les méthodes quantitatives.

Mots-clés : *Modèles ARMA, RCA, Bilinéaires, ARCH, FARMA, Erreur quadratique moyenne de prédiction, Inversibilité, Stationnarité.*

ABSTRACT

In this study, we use mean squared error of prediction to compare the accuracy of various prediction methods for different classes of time series models as ARMA, RCA, BILINEAR, ARCH, ARMA-ARCH and FARMA. This study shows why some methods achieve greater accuracy than others for different types of models. Surprisingly, this study shows that for some time series, easier methods as moving average one perform well in comparison to the more complex and statistically sophisticated methods.

Keywords : *ARMA, RCA, Bilinear, ARCH and FARMA models, Mean squared error of prediction, Invertibility, Stationarity.*

1. Introduction

Etant donné un processus stochastique $\{X(t),\ t \in Z\}$, stationnaire du second ordre, on s'intéresse à calculer sa prédiction à l'horizon h, $h > 0$ à partir de plusieurs méthodes et à les comparer. Les prédicteurs considérés sont : le prédicteur naïf, le prédicteur moyenne mobile simple, le prédicteur lissage exponentiel simple et le prédicteur espérance conditionnelle par rapport aux valeurs passées. Pour comparer ces méthodes, on utilise le critère de l'erreur quadratique moyenne de prédiction

(EQMP) définie par :

$$\sigma^2(e_t(h)) = E[X(t+h) - X_t^*(h)]^2, \quad h > 0$$

où $X_t^*(h)$ représente le prédicteur obtenu à partir de l'une des méthodes précitées.

Dans la littérature, de nombreuses études ont été effectuées sur la comparaison de ces méthodes : on peut consulter par exemple et sans vouloir être exhaustif les travaux de Gross et Ray (1965), Kirby (1966), Lévine (1967), Raine (1971), Reid (1971), Krampf (1972), Groff (1973), Newbold et Granger (1974), Geurts et Ibrahim (1975), Mabert (1975), Makridakis et Hibon (1979), Makridakis *et al* (1984). On peut consulter Bosq et Lecoutre (1992, Chapitre 8.1) pour une revue et une discussion des nombreuses études comparatives des méthodes de prévision. Notons que toutes ces études comparatives ont été effectuées sur une collection de séries chronologiques diverses. Notre étude diffère des études précédentes dans le sens où on s'intéresse à la comparaison des différentes méthodes sur des modèles théoriques et non sur des données réelles. Les modèles théoriques retenus pour notre étude appartiennent aux classes de modèles telles que ARMA (Box et Jenkins, 1970), RCA (Nicholls et Quinn, 1982), Bilinéaire (Granger et Anderson, 1978), ARCH (Engle, 1982), ARMA – ARCH (Weiss, 1984), FARMA (Granger et Joyeux 1980; Hoskins, 1981). Notre approche est similaire à celle de Carbon et Delecroix (1993) qui avaient comparé la méthode non paramétrique et la méthode de Box-Jenkins à partir des critères EMO (erreur relative moyenne observée) et EMP (erreur relative moyenne de prévision). Le principal objet de notre étude est de comparer les prédicteurs considérés sur ces différentes classes de modèles.

Notre article est organisé de la manière suivante : Dans le paragraphe 2, on définit les différents prédicteurs. Dans le paragraphe 3, l'erreur quadratique moyenne de prédiction $\sigma^2(e_t(h))$ est calculée pour chacune de ces méthodes dans le cas d'un processus stationnaire du second ordre inversible. Dans le paragraphe 4, nous appliquons ces résultats à un certain nombre de modèles. Enfin nous présentons quelques simulations dans le paragraphe 5. Le paragraphe 6 donne les conclusions de cet article et dans le paragraphe 7 nous donnons quelques démonstrations.

2. Définitions des prédicteurs

On observe une série jusqu'à l'instant t et on cherche à prédire à partir de différentes méthodes la valeur qu'elle prendra à l'instant $t+h$, $h \in N^*$. Nous présentons ici les différents prédicteurs.

– Le prédicteur naïf (NAF) : il est donné par la dernière observation :

$$X_t^*(h) = X(t) \tag{2.1}$$

– Le prédicteur Moyenne Mobile Simple (MBS) de longueur $m+1$ est la moyenne arithmétique :

$$X_t^*(h) = \frac{1}{m+1}[X(t) + X(t-1) + X(t-2) + \ldots + X(t-m)] \tag{2.2}$$

où m est précisé dans la suite.

– Le prédicteur Lissage Exponentiel Simple (LES) défini par :

$$X_t^*(h) = \alpha \sum_{i=0}^{\infty} (1-\alpha)^i X(t-i) \tag{2.3}$$

où $\alpha (0 < \alpha < 1)$ est la constante de lissage.

– Le prédicteur Espérance Conditionnelle (ECO) défini par

$$X_t^*(h) = E[X(t+h)/F(t)] \tag{2.4}$$

où $F(t) = \sigma(X(s),\ s \leqslant t), \sigma(.)$ désigne la tribu engendrée par le passé et le présent du processus $X(t)$. Le prédicteur ECO est un prédicteur probabiliste, il est par définition celui qui minimise l'erreur quadratique moyenne de prédiction. Il est tout simplement utilisé comme étalon. Les autres prédicteurs sont des prédicteurs statistiques, ils sont construits à partir des observations. Les différentes méthodes ci-dessus sont toutes fondées sur la notion de pondération qu'on affecte à des observations récentes de l'évènement à prévoir; elles ne sont pas nécessairement aussi différentes qu'il peut paraître. Pour l'établir, il faut penser en termes de modèle sous-jacent. Rappelons que Muth (1960) a été le premier auteur parmi tant d'autres à établir un lien entre le lissage exponentiel et le modèle ARIMA. Il montra que la formule du lissage exponentiel simple est la fonction de prévision optimale du modèle ARIMA (0, 1, 1). Notons que le prédicteur naïf est optimal pour un processus de marche aléatoire (ou martingale) c'est-à-dire tel que : $E[X(t+h)/F(t)] = X(t)$, pour toute date t et tout horizon h, $h > 0$. Finalement, remarquons que le prédicteur moyenne mobile simple d'ordre $m+1$ est la fonction de prévision optimale du modèle ARIMA $(m, 1, 0)$.

3. Erreur quadratique moyenne de prédiction

Nous considérons ici un processus $\{X(t),\ t \in Z\}$ stationnaire du second ordre, inversible, centré et de fonction d'autocovariance $\gamma(h) = E[X(t)X(t+h)]$. Dans ce qui suit nous donnons l'expression de l'erreur quadratique moyenne de prédiction à l'horizon h $\sigma^2(e_t(h)) = E[X(t+h) - X_t^*(h)]^2$ associée aux différents prédicteurs définis dans le paragraphe précédent pour ce processus.

– Pour le prédicteur naïf on a :

$$\sigma^2(e_t(h)) = 2[\gamma(0) - \gamma(h)] \tag{3.1}$$

– Pour le prédicteur moyenne mobile simple on obtient :

$$\sigma^2(e_t(h)) = \frac{m+2}{m+1}\gamma(0) - \frac{2}{m+1}\sum_{i=0}^{m}\gamma(i+h) + \frac{2}{(m+1)^2}\sum_{l=1}^{m}(m-l+1)\gamma(l) \tag{3.2}$$

Le calcul se trouve dans l'Appendice 1.

– Pour le prédicteur lissage exponentiel simple on a :

$$\sigma^2(e_t(h)) = \frac{2}{2-\alpha}\gamma(0) - 2\alpha\sum_{i=0}^{\infty}(1-\alpha)^i\gamma(i+h) + \frac{2\alpha}{2-\alpha}\sum_{l=1}^{\infty}(1-\alpha)^l\gamma(l) \quad (3.3)$$

Le calcul se trouve dans l'Appendice 2.

– Pour faire le calcul dans le cas du prédicteur espérance conditionnelle (ou optimal), on considère la représentation moyenne mobile infinie du processus $\{X(t)\}$, à savoir :

$$X(t) = \sum_{i=0}^{\infty}\psi_i\varepsilon(t-i)$$

où les poids ψ_i définissant le processus sont des nombres réels qui vérifient : $\sum_{i=0}^{\infty}|\psi_i| < +\infty$, avec $\psi_0 = 1$ et où $\varepsilon(t)$ est un bruit blanc de variance σ^2, alors on a :

$$\sigma^2(e_t(h)) = \sigma^2\left(1 + \sum_{i=0}^{h-1}\psi_i^2\right) \quad (3.4)$$

Dans les formules (3.1)-(3.3), notons que l'expression de l'erreur quadratique est obtenue en fonction de l'autocovariance $\gamma(h)$. Si le processus considéré est non centré d'espérance μ, alors il suffira de remplacer dans les formules ci-dessus $\gamma(h)$ par $c(h) + \mu^2$ où $c(h)$ est la fonction d'autocovariance du processus centré. L'approche est complètement différente pour le prédicteur optimal.

4. Comparaisons des erreurs quadratiques

Nous étudions essentiellement les prédicteurs pour les processus ARMA(p,q) et ARCH(p). Nous appliquons ensuite ces résultats aux processus RCA et Bilinéaire. Nous utilisons le critère de minimalité de l'erreur quadratique pour comparer les performances des prédicteurs NAF, MBS et LES. Dans tout ce qui suit le prédicteur ECO considéré est simplement utilisé comme étalon.

4.1 Modèles ARMA

Les modèles ARMA ont été introduits par Box et Jenkins en 1970. On dit que le processus $\{X(t),\ t \in Z\}$ suit un modèle ARMA(p,q), s'il est stationnaire du second ordre et défini par la relation suivante :

$$\Phi(B)X(t) = \Theta(B)\varepsilon(t), \quad t \in Z \quad (4.1)$$

$\{\varepsilon(t),\ t \in Z\}$ est un bruit blanc de variance σ^2 tel que :

$$E(\varepsilon_t X_{t^*}) = 0, \quad t^* \leqslant t-1$$

Φ et Θ sont des polynômes à coefficients réels de degrés p et q respectivement et sont définis par :

$$\Phi(z) = 1 - \phi_1 z - \ldots - \phi_p z^p, \quad \phi_p \neq 0$$
$$\Theta(z) = 1 + \theta_1 z + \ldots + \theta_q z^q, \quad \theta_q \neq 0$$

B est l'opérateur retard défini par :

$$B^k X(t) = X(t-k), \quad k \in Z.$$

Le processus défini en (4.1) est stationnaire et inversible si les racines des polynômes Φ et Θ sont de modules strictement supérieurs à un. Ce processus admet la représentation moyenne mobile infinie suivante

$$X(t) = \Psi(B)\varepsilon(t) = \sum_{i=0}^{\infty} \psi_i \varepsilon(t-i)$$

où les poids ψ_i sont définis par $\psi_0 = 1, \psi_i = \theta_i + \sum_{k=1}^{\min(i,p)} \phi_k \psi_{i-k}, i = 1, 2, \ldots$ et par convention $\theta_i = 0$ pour $i > q$ et $\phi_i = 0$ pour $i > p$. Sa fonction d'autocovariance est alors $\gamma(h) = \sigma^2 \sum_{i=0}^{\infty} \psi_i \psi_{i+h}$. Nous allons comparer les performances des différents prédicteurs sur les deux processus particuliers : AR(1) et MA(1). On notera par $EQi(h)$ l'erreur quadratique moyenne de prédiction à l'horizon h où i représente, dans la suite, le numéro du prédicteur utilisé. Pour des raisons de commodité nous considérons les prédicteurs dans l'ordre suivant : ECO, LES, MBS et NAF.

– Soit le processus $(X(t))$ défini par $X(t) = \phi X(t-1) + \varepsilon(t)$, $|\phi| < 1$. Pour ce processus $\psi_i = \phi^i$, $i \geqslant 0$. Ce processus est centré, de variance $\gamma(0) = \sigma^2(1-\phi^2)^{-1}$ et de fonction d'autocovariance $\gamma(h) = \gamma(0)\phi^h$. Posons : $a = 1 - (1-\alpha)\phi$ et $b = 2 - \alpha$. Pour ce processus on a les résultats suivants pour les prédicteurs : ECO, LES, MBS et NAF.

(i) ECO : $EQ1(h) = (1 - \phi^{2h})\gamma(0)$

(ii) LES : $EQ2(h) = \left(\frac{2}{b} + \frac{2(1-\alpha)\phi\alpha}{ab} - \frac{2\alpha}{a}\phi^h\right)\gamma(0)$

Un choix conseillé de la constante de lissage est $\alpha = 0.3$ (Voir à ce sujet Cox (1961) ou Brown (1962), Chapitre 8). Alors, dans ce cas

$$EQ2(h) = \left(1.1764 + 0.247\frac{\phi}{1-0.7\phi} - 0.6\frac{\phi^h}{1-0.7\phi}\right)\gamma(0)$$

(iii) MBS :

$$EQ3(h) = \left(\frac{m+2}{m+1} - \frac{2(1-\phi^{m+1})}{(m+1)(1-\phi)}\phi^h\right)\gamma(0)$$
$$+ \frac{2}{(m+1)^2}\left(\frac{(m+1)(1-\phi^m)\phi}{1-\phi} - \frac{1-[1+(1-\phi)m]\phi^m}{(1-\phi)^2}\phi\right)\gamma(0)$$

(iv) NAF : $EQ4(h) = 2(1-\phi^h)\gamma(0)$

Nous avons calculé les valeurs des différentes erreurs quadratiques pour $h = 1, \ldots, 10$ et ceci pour les valeurs de ϕ comprises entre 0,1 et 0,9 avec un pas de 0,1. Nous donnons les valeurs de $EQi(h)$ en fonction de $\gamma(0)$ pour $h = 1$ et $i = 1, \ldots, 4$ dans le tableau 1.

TABLEAU 1
Erreur quadratique moyenne de prédiction à l'horizon $h = 1$
des différents prédicteurs sur le modèle
$X(t) = \phi X(t-1) + \varepsilon(t)$

Valeurs de ϕ	ECO	LES ($\alpha = 0.3$)	MBS ($m = 10$)	MBS ($m = 20$)	MBS ($m = 50$)	NAF
0.1	**0.9900**	1.1393	1.0888	1.0470	1.0195*	1.8000
0.2	**0.9600**	1.0986	1.0857	1.0462	1.0193*	1.6000
0.3	**0.9100**	1.0544	1.0807	1.0448	1.0191*	1.4000
0.4	**0.8400**	1.0078*	1.0725	1.0425	1.0187	1.2000
0.5	**0.7500**	0.9619*	1.0579	1.0385	1.0180	1.0000
0.6	**0.6400**	0.9262	1.0301	1.0306	1.0167	0.8000*
0.7	**0.5100**	0.9292	0.9732	1.0124	1.0136	0.6000*
0.8	**0.3600**	1.0735	0.8511	0.9621	1.0042	0.4000*
0.9	**0.1900**	1.9400	0.5836	0.7779	0.9523	0.2000*

Dans le tableau 1 on a noté en gras et par un * les deux plus petites erreurs quadratiques. Au vu des résultats on constate qu'à l'horizon de prédiction $h = 1$, le prédicteur naïf est meilleur que les prédicteurs LES et MBS si $0.6 \leqslant \phi \leqslant 0.9$. Le prédicteur LES est meilleur que le prédicteur MBS si $0.4 \leqslant \phi \leqslant 0.6$. Notons que le prédicteur MBS d'ordre élevé a de bonnes performances prédictives quand l'horizon de prédiction s'éloigne.

– Soit le processus $X(t)$, défini par : $X(t) = \varepsilon(t) + \theta\varepsilon(t-1)$, $|\theta| < 1$. Pour ce processus $\psi_0 = 1$, $\psi_1 = \theta$ et $\psi_i = 0$ pour $i \geqslant 2$. Ce processus est centré, de variance $\gamma(0) = (1+\theta^2)\sigma^2$ et de fonction d'autocovariance $\gamma(h) = \sigma^2\theta$, si $h = 1$ et $\gamma(h) = 0$, si $h > 1$. Posons $\alpha = 0.10$ et $\sigma^2 = 1$. Pour ce processus on a les résultats suivants pour les prédicteurs ECO, LES, MBS et NAF

$$
\begin{aligned}
\text{(i) : ECO} \quad EQ1(h) &= 1 \text{ si } h = 1 \\
EQ1(h) &= \gamma(0) \\
&= 1 + \theta^2, \text{ si } h > 1 \\
\text{(ii) : LES} \quad EQ2(h) &= \frac{2}{2-\alpha}\gamma(0) - 2\alpha\gamma(1) + \frac{2(1-\alpha)\alpha}{2-\alpha}\gamma(1) \\
&= 1.052(1+\theta^2) - 0.105\,\theta, \text{ si } h = 1 \\
EQ2(h) &= \frac{2}{2-\alpha}\gamma(0) + \frac{2(1-\alpha)\alpha}{2-\alpha}\gamma(1) \\
&= 1.052(1+\theta^2) + 0.095\,\theta, \text{ si } h > 1 \\
\text{(iii) : MBS} \quad EQ3(h) &= \frac{m+2}{m+1}\gamma(0) - \frac{2}{m+1}\gamma(1) + \frac{2m}{(m+1)^2}\gamma(1) \\
&= \frac{m+2}{m+1}(1+\theta^2) - \frac{2}{(m+1)^2}\theta, \text{ si } h = 1 \\
EQ3(h) &= \frac{m+2}{m+1}\gamma(0) + \frac{2m}{(m+1)^2}\gamma(1) \\
&= \frac{m+2}{m+1}(1+\theta^2) + \frac{2m}{(m+1)^2}\theta, \text{ si } h > 1 \\
\text{(iv) : NAF} \quad EQ4(h) &= 2[\gamma(0) - \gamma(1)] \\
&= 2(1+\theta^2-\theta), \text{ si } h = 1 \\
EQ4(h) &= 2\gamma(0) \\
&= 2(1+\theta^2), \text{ si } h > 1
\end{aligned}
$$

Nous avons calculé les valeurs des différentes erreurs quadratiques pour $h = 1$ et pour $h > 1$ et pour les valeurs de θ comprises entre 0.1 et 0.9 avec un pas de 0.1. Nous donnons les valeurs de $EQi(h)$ dans le tableau 2, les valeurs supérieures correspondent à l'horizon $h = 1$ et les valeurs inférieures à l'horizon $h > 1$.

Au vu des résultats du tableau 2 on constate qu'à l'horizon $h = 1$, le prédicteur LES est meilleur que le prédicteur MBS si $0.4 \leqslant \theta \leqslant 0.9$. On note aussi que le prédicteur MBS a des performances prédictives supérieures à celles des prédicteurs LES et NAF quand $h > 1$.

4.2 Modèles ARCH

Les modèles ARCH ont été introduits par Engle en 1982. On dit que le processus $\{X(t),\ t \in Z\}$ suit un processus ARCH(p) s'il est défini par

$$X(t) = \varepsilon(t)\sqrt{\lambda_0 + \sum_{i=1}^{p} \lambda_i [X^2(t-i) - \lambda_0]}$$

TABLEAU 2
Erreur quadratique moyenne de prédiction des différents prédicteurs sur le modèle $X(t) = \varepsilon(t) + \theta\varepsilon(t-1)$

Valeurs de θ	ECO	LES ($\alpha = 0.10$)	MBS ($m = 10$)	MBS ($m = 20$)	MBS ($m = 50$)	NAF
0.1	**1.00**	1.05	1.10	1.05	1.02*	1.82
	1.01	1.07	1.12	1.06	1.03*	2.02
0.2	**1.00**	1.07	1.13	1.08	1.06*	1.68
	1.04	1.11	1.16	1.10	1.07*	2.08
0.3	**1.00**	1.11	1.18	1.14	1.11*	1.58
	1.09	1.17	1.23	1.17	1.12*	2.18
0.4	**1.00**	1.17*	1.25	1.21	1.18	1.52
	1.16	1.25	1.33	1.25	1.19*	2.32
0.5	**1.00**	1.26*	1.35	1.30	1.27	1.50
	1.25	1.36	1.44	1.35	1.29*	2.50
0.6	**1.00**	1.36*	1.47	1.42	1.38	1.52
	1.36	1.48	1.58	1.48	1.40*	2.72
0.7	**1.00**	1.49*	1.61	1.55	1.51*	1.58
	1.49	1.63	1.74	1.62	1.54*	2.98
0.8	**1.00**	1.64*	1.77	1.71	1.67	1.68
	1.64	1.80	1.92	1.79	1.70*	3.28
0.9	**1.00**	1.81*	1.95	1.89	1.84	1.82
	1.81	1.98	2.12	1.97	1.88*	3.62

où $\varepsilon(t)$ est un bruit blanc gaussien. Sous les conditions de régularité : $\lambda_0 > 0$, $\lambda_i \geqslant 0$, $i = 1, \ldots, p$ et $\sum_{i=1}^{p} \lambda_i < 1$, Milhoj (1985) a montré que le processus ARCH(p) est non corrélé, de moyenne nulle et de variance λ_0 : $E[X(t)] = 0, \gamma(0) = \lambda_0, \gamma(h) = 0$ pour $h > 0$. L'erreur quadratique associée au prédicteur de l'espérance conditionnelle vaut $\sigma^2(e_t(h)) = \lambda_0$. Pour ce modèle on a les résultats suivants pour les différents prédicteurs

(i) : ECO $\quad EQ1(h) = \gamma(0)$

(ii) : LES $\quad EQ2(h) = \dfrac{2}{2-\alpha}\gamma(0)$. Pour $\alpha = 0.10$ on a

$EQ2(h) = 1.0526\gamma(0)$

(iii) : MBS $\quad EQ3(h) = \dfrac{m+2}{m+1}\gamma(0)$.

On remarque que $EQ3(h) = 1.0526\gamma(0)$ si $m = 18$

(iv) : NAF $\quad EQ4(h) = 2\gamma(0)$

En utilisant le critère de minimalité de l'erreur quadratique moyenne de prédiction on a les conclusions suivantes quand on prend $\alpha = 0.10$:

– le prédicteur LES est meilleur que le prédicteur MBS si $m < 18$
– les prédicteurs LES et MBS ont des performances similaires si $m = 18$
– le prédicteur MBS est meilleur que le prédicteur LES si $m > 18$
– les prédicteurs LES et MBS sont meilleurs que le prédicteurs NAF

D'où le classement suivant

Pour $m < 18$	Pour $m > 18$
1- Lissage Exponentiel Simple	1- Moyenne Mobile Simple
2- Moyenne Mobile Simple	2- Lissage Exponentiel Simple
3- Naïf	3- Naïf

Nous faisons maintenant quelques remarques concernant des modèles non linéaires ayant des comportements similaires aux modèles AR(p) ou ARCH(p).

1) Soit le processus RCA(1) (Nicholls et Quinn, 1982) défini par :

$$X(t) = (a + b(t))X(t-1) + \varepsilon(t)$$

où a est un paramètre fixé, $\varepsilon(t)$ est un bruit blanc de variance 1, $b(t)$ est une variable aléatoire telle que : $E(b(t)) = 0$, $E(b^2(t)) = \lambda$, $b(t)$ est indépendante de $X(t)$ et de $\varepsilon(t)$. On suppose que $a^2 + \lambda < 1$. Le processus est centré et sa variance est $\gamma(0) = (1 - a^2 - \lambda)^{-1}$. L'erreur quadratique associée au prédicteur ECO est $\sigma^2(e_t(h)) = \gamma(0)(1 - a^{2h})$, (Ray, 1983). On constate que $\gamma(h) = \gamma(0)a^h$ a la même structure que celle d'un AR(1) donc les performances des différents prédicteurs sur le modèle RCA(1) sont similaires à celles obtenues pour le processus autorégressif d'ordre un.

2) Considérons le modèle bilinéaire d'ordre 1 (Granger et Anderson, 1978), défini par

$$X(t) = bX(t-2)\varepsilon(t-1) + \varepsilon(t)$$

où les $\varepsilon(t)$ sont i.i.d de moyenne nulle et de variance 1, $\varepsilon(t)$ est indépendante de $X(s)$, $s < t$. Le processus est stationnaire si $|b| < 1$, il est inversible si $2b^2 < 1$. De plus le processus est centré et sa variance est $\gamma(0) = 1/(1 - b^2)$. L'erreur quadratique associée au prédicteur de l'espérance conditionnelle est : $\sigma^2(e_t(h)) = 1$ si $h = 1$ et $\sigma^2(e_t(h)) = \gamma(0)$, si $h > 1$. On constate que $\gamma(h) = 0$, quelque soit $h \neq 0$. La fonction d'autocovariance du modèle bilinéaire superdiagonal d'ordre un a donc la même structure que celle d'un ARCH(p). Les performances des différents prédicteurs sur ce modèle sont les mêmes que celles obtenues pour le modèle ARCH(p).

3) Soit le modèle AR(1) avec erreur ARCH(1) introduit par Weiss (1984) défini par

$$X(t) = \phi X(t-1) + \varepsilon(t)\sqrt{\lambda_0 + \lambda_1 X^2(t-1)},\ |\phi| < 1,\ \lambda_0 > 0,\ \lambda_1 < 1$$

où les $\varepsilon(t)$ sont des variables aléatoires i.i.d de moyenne nulle et de variance 1. La variance du processus est $\gamma(0) = \dfrac{\lambda_0}{(1 - \phi^2)(1 - \lambda_1)}$. L'erreur quadratique associée au prédicteur ECO est $\sigma^2(e_t(h)) = \gamma(0)(1-\phi^{2h})$, (Guégan, 1992). On constate alors que $\gamma(h) = \gamma(0)\cdot\phi^h$ a la même structure que celle d'un AR(1), donc les performances des différents prédicteurs sur ce modèle sont les mêmes que celles obtenues pour le processus autorégresssif d'ordre un.

5. Simulations

Dans ce paragraphe nous donnons quelques simulations en considérant les prédicteurs NAF, MBS, LES et ECO. Notre intention est de mettre en évidence la meilleure méthode pour prédire à un horizon donné. Pour tous les modèles simulés, nous considérons une taille d'échantillon de 100 observations. Nous utilisons le logiciel Maple V pour le calcul des erreurs quadratiques. Dans le tableau 3, nous indiquons les modèles simulés.

TABLEAU 3

Modèles simulés	
1 AR(1)	$X(t) = 0,3X(t-1) + \varepsilon(t)$
2 MA(2)	$X(t) = \varepsilon(t) - 0,65\varepsilon(t-1) - 0,24\varepsilon(t-2)$
3 RCA(1)	$X(t) = (0,5 + b(t))X(t-1) + \varepsilon(t)$ où $b(t) \sim \text{MA}(1)$
	$b(t) = e(t) - 0,5e(t-1)$
4 BLSD(1)	$X(t) = 0,25X(t-2)\varepsilon(t-1) + \varepsilon(t)$
5 AR(1)–ARCH(1)	$X(t) = 0,45X(t-1) + \varepsilon(t)\sqrt{0,35 + 0,25X^2(t-1)}$
6 ARMA(2, 1)	$X(t) = X(t-1) - 0,25X(t-2) + \varepsilon(t) + \varepsilon(t-1)$
7 FARMA(0, 0,20, 0)	$(1-B)^{0,20}X(t) = \varepsilon(t)$
8 BL(1,0,1,1)	$X(t) = 0,45X(t-1) + 0,25X(t-1)\varepsilon(t-1) + \varepsilon(t)$

Pour les modèles ci-dessus, $\varepsilon(t)$ est un bruit blanc gaussien réduit et $e(t)$ un bruit blanc gaussien de variance 0,25. Pour le prédicteur du lissage exponentiel simple, nous avons choisi la constante de lissage α qui minimise l'erreur quadratique et pour celui de la moyenne mobile simple différents choix de l'ordre m sont considérés : $m = 10$, $m = 20$ et $m = 50$. Dans les différents tableaux qui suivent, nous donnons les résultats complets concernant ces modèles. Dans tous ces tableaux, nous notons en gras et par un * les deux plus petites erreurs quadratiques.

Les résultats obtenus dans les tableaux 4 à 8 correspondent bien aux résultats théoriques précédemment établis. On note que le prédicteur de la moyenne mobile simple d'ordre élevé a des performances prédictives supérieures au prédicteur du lissage exponentiel simple.

Nous nous intéressons maintenant aux comportements des prédicteurs sur trois processus pour lesquels nous n'avons pas établi de résultats théoriques, à savoir les processus ARMA(2, 1), FARMA(0, d, 0) [Granger et Joyeux (1980), Hoskins (1981)]

TABLEAU 4

Modèle 1 : AR(1) Erreur quadratique moyenne de prédiction						
Lead time (h)	ECO	NAF	LES ($\alpha = 0.30$)	MBS ($m = 10$)	MBS ($m = 20$)	MBS ($m = 50$)
1	**1.0000**	1.5384	1.1452	1.1876	1.1481	1.1199*
2	**1.0900**	2.0000	1.3204	1.2476	1.1795	1.1328*
3	**1.0981**	2.1384	1.3730	1.2656	1.1889	1.1367*
4	**1.0988**	2.1800	1.3888	1.2709	1.1918	1.1379*
5	**1.0988**	2.1924	1.3935	1.2726	1.1926	1.1382*

TABLEAU 5

Modèle 2 : MA(2) Erreur quadratique moyenne de prédiction						
Lead time (h)	ECO	NAF	LES ($\alpha = 0.10$)	MBS ($m = 10$)	MBS ($m = 20$)	MBS ($m = 50$)
$h = 1$	**1.0000**	3.9482	1.6327	1.3154*	1.3898	1.4418
$h = 2$	**1.4225**	3.4402	1.5387	1.5409	1.5079	1.4904*
$h > 2$	**1.4801**	2.9602	1.4809	1.4972	1.4850	1.4810*

TABLEAU 6

Modèle 3 : RCA(1) Erreur quadratique moyenne de prédiction						
Lead time (h)	ECO	NAF	LES ($\alpha = 0.30$)	MBS ($m = 10$)	MBS ($m = 20$)	MBS ($m = 50$)
1	**1.7142**	2.2857	2.0684*	2.4118	2.3738	2.3270
2	**2.1428**	3.4285	2.5958	2.6264	2.4826	2.3718*
3	**2.2500**	4.0000	2.8596	2.7302	2.5370	2.3942*
4	**2.2767**	4.2857	2.9914	2.7822	2.5643	2.4054*
5	**2.2834**	4.4285	3.0574	2.8081	2.5779	2.4110*

TABLEAU 7

Modèle 4 : BLSD(1) Erreur quadratique moyenne de prédiction						
Lead time (h)	ECO	NAF	LES ($\alpha = 0.10$)	MBS ($m = 10$)	MBS ($m = 20$)	MBS ($m = 50$)
$h = 1$	**1.0000**	2.1332	1.1227	1.1635	1.1173	1.0875*
$h > 1$	**1.0666**	2.1332	1.1227	1.1635	1.1173	1.0875*

TABLEAU 8

Modèle 5 : AR(1)-ARCH(1) Erreur quadratique moyenne de prédiction						
Lead time (h)	ECO	NAF	LES ($\alpha = 0.30$)	MBS ($m = 10$)	MBS ($m = 20$)	MBS ($m = 50$)
1	**0.4666**	0.6436	0.5527*	0.6240	0.6090	0.5959
2	**0.5611**	0.9333	0.6795	0.6718	0.6341	0.6062*
3	**0.5803**	1.0636	0.7366	0.6934	0.6454	0.6109*
4	**0.5841**	1.1223	0.7623	0.7031	0.6505	0.6130*
5	**0.5849**	1.1487	0.7739	0.7074	0.6528	0.6139*

et BL(1,0,1,1). Le processus FARMA(0, d, 0) est défini par :

$$(1 - B)^d X(t) = \varepsilon(t), \quad -1/2 < d < 1/2$$

où les variables aléatoires $\varepsilon(t)$ sont indépendantes, identiquement distribuées gaussiennes centrées et de variance 1. Ce processus est à mémoire longue, sa fonction d'autocorrélation est telle :

$$\rho(h) \sim ch^{2d-1}, \text{ quand } h \to \infty, \ c > 0, \ d < 1/2,$$

où $\sim$ représente l'équivalence asymptotique. Notons aussi que tous les modèles décrits précédemment sont à mémoire courte, leur fonction d'autocorrélation est géométriquement bornée. Le modèle FARMA prend bien en compte les prévisions à long terme. Le modèle bilinéaire diagonal d'ordre un, noté BL(1, 0, 1, 1) est défini par :

$$X(t) = aX(t-1) + bX(t-1)\varepsilon(t-1) + \varepsilon(t)$$

où $\varepsilon(t)$ défini comme précédemment est indépendant de $X(s)$, $s < t$. On suppose que $|a| < 1$ et $a^2 + b^2 < 1$. Pour calculer les erreurs quadratiques associées à chaque prédicteur, nous avons utilisé pour ces processus les formules (3.1)-(3.4) en remplaçant $\gamma(0)$, $\gamma(h)$ et ψ_i par leurs expressions. Pour le processus ARMA(2, 1), ces expressions sont :

$$\gamma(0) = \frac{32}{3},\ \gamma(h) = 2^{-h}\left(\frac{32}{3} + 8h\right) \text{ et } \psi_i = (1+3i)\cdot 2^{-i},\ i = 0,1,2,\ldots$$

(Brockwell et Davis, 1991). Pour le processus FARMA(0, d, 0) on obtient :

$$\gamma(h) = \frac{(-1)^h\Gamma(1-2d)}{\Gamma(h-d+1)\Gamma(1-h-d)} \text{ et } \psi_i = \frac{\Gamma(d+i)}{\Gamma(d)\Gamma(i+1)},\ i = 0,1,2,\ldots$$

et $\Gamma(.)$ est la fonction gamma. Pour le processus BL(1, 0, 1, 1), on a les résultats suivants, (Voir Sesay et Subba Rao, 1988) :

$$\mu = E(X(t)) = \frac{b}{1-a},\ \gamma(0) = E(X^2(t)) = \frac{1+2b+4ab\mu}{1-a^2-b^2},$$
$$\gamma(1) = a\gamma(0) + 2b\mu \quad \text{et} \quad \gamma(h) = a\gamma(h-1) + b\mu \text{ si} h \geqslant 2$$

Pour calculer l'erreur quadratique associée au prédicteur ECO, on remarque que le processus BL(1, 0, 1, 1) admet la représentation markovienne suivante, (voir Pham, 1985) :

$$X(t) = Z(t-1) + \varepsilon(t)$$
$$Z(t) = (a + b\varepsilon(t))Z(t-1) + (a + b\varepsilon(t))\varepsilon(t)$$

D'où en utilisant un résultat de Guégan (1992), on obtient

$$\sigma^2(e_t(h)) = 1 + \frac{(a^2+3b^2)[1-(a^2+b^2)^{h+1}]}{1-a^2-b^2}$$

Nous donnons les valeurs de $EQi(h)$ pour les processus ARMA(2, 1), FARMA(0, d, 0) et BL(1,0,1,1) dans les tableaux 9, 10 et 11.

TABLEAU 9

Modèle 6 : ARMA(2, 1) Erreur quadratique moyenne de prédiction						
Lead time (h)	ECO	NAF	LES ($\alpha = 0.10$)	MBS ($m = 10$)	MBS ($m = 20$)	MBS ($m = 50$)
1	**1.0000**	2.6700*	8.8153	10.5066	10.8610	10.8257
2	**5.0000**	8.0050*	10.3804	12.1990	11.7497	11.1917
3	**8.0625**	12.6725	11.5265	13.4086	12.3844	11.4532*
4	**9.6250**	16.0062	12.2814	14.1951	12.7970	11.6231*
5	**10.2851**	18.1731	12.7498	14.6792	13.0509	11.7277*

Au vu des résultats, on constate qu'à l'horizon très court terme ($h = 1$ et $h = 2$), le prédicteur naïf est meilleur que les prédicteurs LES et MBS, et le prédicteur LES est meilleur à son tour que le prédicteur MBS. D'autre part on note que le prédicteur MBS est meilleur que les prédicteurs NAF et LES quand l'horizon s'éloigne ($h \geqslant 3$).

TABLEAU 10

Modèle 7 : FARMA(0, 0.20, 0) Erreur quadratique moyenne de prédiction						
Lead time (h)	ECO	NAF	LES ($\alpha = 0.10$)	MBS ($m = 10$)	MBS ($m = 20$)	MBS ($m = 50$)
1	**1.0000**	1.6480	1.1565	1.0993	1.0851	1.0821*
2	1.2000	1.8311	1.1933*	1.1378	1.1071	**1.0919**
3	1.3200	1.9096	1.2139*	1.1603	1.1205	**1.0980**
4	1.4080	1.9550	1.2281*	1.1760	1.1303	**1.1027**
5	1.4784	1.9853	1.2388*	1.1881	1.1380	**1.1064**
10	1.7122	2.0574	1.2701*	1.2236	1.1621	**1.1190**
15	1.8619	2.0876	1.2865*	1.2424	1.1759	**1.1268**
20	1.9748	2.1050	1.2972*	1.2545	1.1852	**1.1325**

Au vu des résultats, on constate que le prédicteur de la moyenne mobile simple apparaît comme la meilleure méthode de prédiction. Cette anomalie est certainement due à l'utilisation de la fonction gamma dans le calcul des erreurs quadratiques.

TABLEAU 11

Modèle 8 : BL(1, 0, 1, 1) Erreur quadratique moyenne de prédiction						
Lead time (h)	ECO	NAF	LES ($\alpha = 0.30$)	MBS ($m = 10$)	MBS ($m = 20$)	MBS ($m = 50$)
1	**0.4933**	2.0964	1.9369*	2.2464	2.1971	2.1512
2	**0.5207**	3.2671	2.4496	2.4399	2.2984	2.1929*
3	**0.5279**	3.7939	2.6803	2.5269	2.3441	2.2117*
4	**0.5299**	4.0310	2.7841	2.5661	2.3646	2.2201*
5	**0.5304**	4.1377	2.8309	2.5837	2.3738	2.2239*

Au vu des résultats, on constate que le prédicteur LES est meilleur que les prédicteurs NAF et MBS à l'horizon $h = 1$. A l'horizon $h \geqslant 2$, on remarque que le prédicteur MBS est meilleur que les prédicteurs NAF et LES.

6. Conclusions

Les résultats présentés dans cette étude indiquent que le prédicteur optimal est bien entendu la meilleure méthode de prédiction. Des résultats analogues ont été obtenus par par Reid (1971) et Newbold et Granger (1974). Ces résultats sont en contradiction avec ceux de Groff (1973) et Geurts et Ibrahim (1975) qui indiquaient que la méthodologie de Box et Jenkins donnait des résultats similaires ou même légèrement inférieurs à ceux du lissage exponentiel. D'autre part si on compare les prédicteurs lissage exponentiel simple et moyenne mobile simple, on note que le prédicteur MBS d'ordre élevé donne pour la plupart des modèles considérés la meilleure erreur quadratique moyenne de prédiction. Cette étude indique aussi que le prédicteur du lissage exponentiel simple est en général supérieur au prédicteur moyenne mobile simple pour la prévision à court terme. Des résultats analogues ont été obtenus par Gross et Ray (1965), Kirby (1966), Lévine (1967), Raine (1971) et Krampf (1972).

Pour conclure, nous indiquons que les études de comparaisons des méthodes de prédiction ne doivent à aucun prix constituer un point de repère pour l'utilisateur dans le choix d'une méthode. Ce choix doit dépendre du secteur d'activité et de l'horizon. Aucune des méthodes n'est définitivement et nettement «meilleure» qu'une autre, chaque méthode a beaucoup d'avantages mais aussi quelques inconvénients. Il y a donc une méthode de prévision optimale et certaines méthodes sont mieux adaptées à certaines situations. Il est aussi à noter que les méthodes quantitatives ne sont pas toujours meilleures que la prévision purement intuitive, ce qui fait que ces deux méthodes ne doivent pas être considérées comme concurrentes mais complémentaires : la prévision formalisée pouvant être un élément qui, comparé à l'intuition, permet de dégager une donnée prévisionnelle. La méthode de prévision «idéale» n'existant pas, certains auteurs préconisent de combiner des méthodes au lieu d'utiliser une seule technique et ont rencontré certains succès, il est cependant difficile de trouver une justification rationnelle à cette procédure.

7. Appendices

1. Erreur quadratique associée au prédicteur de la moyenne mobile simple

L'erreur quadratique associée au prédicteur défini en (2.2) est :

$$\begin{aligned}\sigma^2(e_t(h)) &= E\left(X(t+h) - \frac{1}{m+1}\sum_{i=0}^{m} X(t-i)\right)^2 \\ &= E(X^2(t+h)) - \frac{2}{m+1}\sum_{i=0}^{m} E(X(t+h)X(t-i)) \\ &+ \frac{1}{(m+1)^2}\left(\sum_{i=0}^{m} E(X^2(t-i)) + 2\sum_{i=0}^{m-1}\sum_{j=i+1}^{m} E(X(t-i)X(t-j))\right)\end{aligned}$$

Pour ce processus stationnaire

$$E[X(t)X(t+h)] = \gamma(h)$$

d'où

$$\begin{aligned}\sigma^2(e_t(h)) =&\gamma(0) - \frac{2}{m+1}\sum_{i=0}^{m}\gamma(i+h) + \frac{1}{m+1}\gamma(0)\\ &+ \frac{2}{(m+1)^2}\sum_{l=1}^{m}(m-l+1)\gamma(l)\\ =&\frac{m+2}{m+1}\gamma(0) - \frac{2}{m+1}\sum_{i=0}^{m}\gamma(i+h) + \frac{2}{(m+1)^2}\sum_{l=1}^{m}(m-l+1)\gamma(l)\end{aligned}$$

Et l'on obtient (3.2).

2. Erreur quadratique associée au prédicteur du lissage exponentiel simple

L'erreur quadratique associée au prédicteur défini en (2.3) est :

$$\begin{aligned}\sigma^2(e_t(h)) &= E\left[X(t+h) - \alpha\sum_{i=0}^{\infty}(1-\alpha)^i X(t-i)\right]^2\\ &= E[X^2(t+h)] - 2\alpha\sum_{i=0}^{\infty}(1-\alpha)^i E[X(t+h)X(t-i)]\\ &+ \alpha^2\sum_{i=0}^{\infty}(1-\alpha)^{2i}E[X^2(t-i)]\\ &+ 2\alpha^2\sum_{i=0}^{\infty}\sum_{j=i+1}^{\infty}(1-\alpha)^i(1-\alpha)^j E[X(t-i)X(t-j)]\end{aligned}$$

Pour ce processus stationnaire

$$E[X(t)X(t+h)] = \gamma(h)$$

D'où

$$\begin{aligned}
\sigma^2(e_t(h)) &= \gamma(0) - 2\alpha\sum_{i=0}^{\infty}(1-\alpha)^i\gamma(i+h) + \frac{\alpha^2}{1-(1-\alpha)^2}\gamma(0) \\
&+ \frac{2\alpha^2}{1-(1-\alpha)^2}\sum_{l=1}^{\infty}(1-\alpha)^l\gamma(l) \\
&= \frac{2}{2-\alpha}\gamma(0) - 2\alpha\sum_{i=0}^{\infty}(1-\alpha)^i\gamma(i+h) + \frac{2\alpha}{2-\alpha}\sum_{l=1}^{\infty}(1-\alpha)^l\gamma(l)
\end{aligned}$$

Et l'on obtient (3.3).

Remerciements

Nous tenons à remercier D. Bosq, P. Cazes et D. Guégan pour leurs conseils et suggestions qui nous ont permis d'améliorer ce travail.

Bibliographie

Bosq D. et Lecoutre J.P., (1992). «Analyse et prévision des séries chronologiques», Masson.

Box G.E.P. and Jenkins G.M., (1970). «Time Series Analysis : Forecasting and Control», Holden-Day.

Brockwell P.J. and Davis R.A., (1991). «Time Series : Theory and Methods», 2nd Edn, New York : Springer-Verlag.

Brown R., (1962). «Smoothing, forecasting and prediction», Prentice Hall.

Carbon M. et Delecroix M., (1993). «Non-parametric vs parametric forecasting in time series : a computational point of view», Applied Stochastic Models and Data Analysis, Vol. 9, pp 215, 229.

Cogger K., (1974). «The optimality of general order exponential smoothing», Operational Research, Vol. 22, pp 858, 867.

Cox D., (1961). «Prediction by exponentially weighted moving average and related methods», J.R.S.S. B, Vol. 23, pp 414, 422.

Engle R.F., (1982). «Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation», Econometrica 50, pp 987, 1007.

Geurts M.D. and Ibrahim I.B., (1975). «Comparing the Box Jenkins Approach with the exponentially Smoothed Forecasting Model : Application to Hawaii Tourists», Journal of Marketing Research, Vol. 12, pp 182, 188.

Granger C.W.J. and Anderson A.P., (1978). «An introduction to bilinear time series analysis», Vandenhoeck and Rupecht Gottingen.

Granger C.W.J. and Joyeux R., (1980). «An introduction to long-memory time series models and fractional differencing», J.T.S.A, Vol. 1, pp 15, 29.

Groff G.K., (1973). «Empirical Comparisons of Models for Short-Range Forecasting», Management Science, Vol. 20, pp 22, 31.

Gross D. and Ray J.L., (1965). «A general purpose forecasting simulator», Management Science, Vol. 11, pp 119, 135.

Guégan D., (1981). «Etude d'un modèle non linéaire, le modèle superdiagonal d'ordre 1», CRAS, Série A, t 293, pp 95, 98.

Guégan D., (1992). «On the identification and prediction of non linear models», Proceedings for the Workshop on "New Directions in time series analysis". Minneapolis. July 90, Springer Verlag.

Hoskins J.R.M., (1981). «Fractional differencing», Biometrika, Vol. 68, 1, pp 165, 176.

Kirby R.M., (1966). «A comparison of short and medium range statistical forecasting methods», Management Science, Vol. 4, pp 202, 210.

Krampf R.F., (1972). «The turning point problem in smoothing models», Unpublished Ph. D. Dissertation, University of Cincinnati.

Levine A.H., (1967). «Forecasting Techniques», Management Accounting.

Mabert A., (1975). «An introduction to short term forecasting using the Box Jenkins methodology», American Institute of Industrial Engineers, Atlanta.

Makridakis S. and Hibon M., (1979). «Accuracy of forecasting : an empirical investigation», J. R. S. S. A, Vol. 142, pp 97, 145.

Makridakis S. *et al*, (1984). «The Forecasting Accuracy of Major Time Series Methods», Wiley, Chichester.

Milhoj A., (1985). «The Moment Structure of ARCH process», Scand. J. Statist, Vol. 12, pp 281, 292.

Muth J.F., (1960). «Optimal properties of exponentially weighted forecasts», J.A.S.A, Vol. 55, pp 299, 306.

Newbold P. and Granger C.W.J., (1974). «Experience with forecasting univariate time series and the combination of forecasts», J. R. S. S. A, 137, Part 2, pp 131, 165.

Nicholls D.F. and Quinn B.G., (1982). «Random coefficient autoregressive models. An introduction», Lectures Notes in Statistics. II. Springer Verlag.

Pham T.D., (1985). «Bilinear Markovian representation and bilinear models», Stoch. Processes and their app, Vol 12, pp 295-306.

Pham T.D. and Tran L.T., (1981). «On the first order bilinear time series models», J. Appl. Prob, pp 617-627.

Raine J.E., (1971). «Self-Adaptative Forecasting Considered«, Decision Science.

Ray D., (1983). «On the autoregressive model with random coefficients», Calcutta Statist. Ass. Bull, 32, pp 135, 142.

Reid D.J., (1971). «Forecasting in Action : A Comparison of Forecasting Techniques In Economic Time Series», Proceedings of the Joint Conference of the Operations Research Society, Long-Range Planning and Forecasting.

Sesay S.A.O. and Subba Rao T., (1988). «Yule Walker difference equation for higher order moments and cumulants for bilinear time series models», J.T.S.A, Vol 9, pp 85-401.

Weiss A.A., (1984). «ARMA models with ARCH errors», J.T.S.A, Vol. 3, pp 129, 143.