Estimation de modèles ARMA par la méthode du maximum de vraisemblance *

Jean-Marie Dufour †
Université de Montréal

Première version: Février 1981
Révisions: Février 1991, Septembre 2000
Cette version: 14 mars 2002
Compilé: 14 mars 2002, 11:11am

* Cette recherche a bénéficié du support financier de la Chaire de recherche du Canada en économétrie, du Conseil des Arts du Canada (Bourse Killam), du Conseil de recherche en sciences humaines du Canada, du Conseil de recherche en sciences naturelles et en génie du Canada, du Réseau canadien de centres d'excellence (projet MITACS) et du Fonds FCAR du Québec.

† L'auteur est titulaire de la Chaire de recherche du Canada en économétrie. Centre interuniversitaire de recherche en analyse des organisations (CIRANO), Centre de recherche et développement en économique (C.R.D.E.), et Département de sciences économiques, Université de Montréal. Adresse postale: Département de sciences économiques, Université de Montréal, C.P. 6128 succursale Centre Ville, Montréal, Québec, Canada H3C 3J7. TEL: (514) 343 2400; FAX: (514) 343 5831; courriel: jean.marie.dufour@umontreal.ca. Page Web: http://www.fas.umontreal.ca/SCECO/Dufour .

Table des matières

1. Modèle

Soit

$$X_t \sim ARIMA\,(p,\, d,\, q)\ . \tag{1.1}$$

X_t obéit au modèle :

$$\varphi\,(B)\,\nabla^d X_t = \bar{\mu} + \theta\,(B)\,u_t \tag{1.2}$$

où

$$\begin{aligned}
\nabla &= 1 - B, && (1.3)\\
\varphi\,(B) &= 1 - \varphi_1 B - \cdots - \varphi_p B^p\,,\ \varphi_p \neq 0, && (1.4)\\
\theta\,(B) &= 1 - \theta_1 B - \cdots - \theta_q B^q\,,\ \theta_q \neq 0, && (1.5)\\
& \varphi\,(B) \text{ et } \theta\,(B) \text{ n'ont pas de racine commune,} && (1.6)\\
u_t &\sim BB\left(0,\, \sigma^2\right)\ . && (1.7)
\end{aligned}$$

Série de $N = n + d$ valeurs : $X_{-d+1},\, \ldots\,,\, X_0,\, X_1,\, \ldots\,,\, X_n$ $p+q+2$ paramètres à estimer.

Posons :

$$\varphi = \begin{pmatrix} \varphi_1 \\ \vdots \\ \varphi_p \end{pmatrix}\ ,\quad \theta = \begin{pmatrix} \theta_1 \\ \vdots \\ \theta_q \end{pmatrix} \tag{1.8}$$

$$W_t \equiv \nabla^d X_t\,,\ t = 1,\, \ldots\,,\, n, \tag{1.9}$$

où X_t a possiblement subi une transformation exponentielle ou logarithmique. Nous pouvons écrire :

$$\begin{aligned}
W_t &\sim ARMA\,(p,\, q)\\
\varphi\,(B)\,W_t &= \bar{\mu} + \theta\,(B)\,u_t\\
W_t &= \varphi_1 W_{t-1} + \cdots + \varphi_p W_{t-p}\\
&\quad + u_t - \theta_1 u_{t-1} - \cdots - \theta_q u_{t-q} + \bar{\mu}\\
u_t &= W_t - \varphi_1 W_{t-1} - \cdots - \varphi_p W_{t-p}\\
&\quad + \theta_1 u_{t-1} + \cdots + \theta_q u_{t-q} - \bar{\mu}.
\end{aligned}$$

Si on pose $\tilde{W}_t = W_t - \mu$, où

$$\begin{aligned}
\mu &= \bar{\mu}/\left(1 - \varphi_1 - \cdots - \varphi_p\right),\\
u_t &= \tilde{W}_t - \varphi_1 \tilde{W}_{t-1} - \cdots - \varphi_p \tilde{W}_{t-p} + \theta_1 u_{t-1} + \cdots + \theta_q u_{t-q}\ .
\end{aligned}$$

Supposons que $u_t \overset{i.i.d}{\sim} N\,(0,\, \sigma^2)\,.$

Difficulté : à $t = 1,\ W_{t-1},\ \ldots\ ,\ W_{t-p},\ u_{t-1},\ \ldots\ ,\ u_{t-q}$ sont inconnus.
On peut ici utiliser deux approches principales :

1. maximiser la fonction de vraisemblance conditionnelle ;
2. maximiser la fonction de vraisemblance inconditionnelle.

2. Fonction de vraisemblance conditionnelle

Soit

$$\begin{array}{l} W_* : p \text{ observations de } W_t \text{ antérieures au début de la série}, \\ u_* \ : q \text{ bruits } u_t, \\ W_* = (W_0,\ W_{-1},\ \ldots\ ,\ W_{-p+1})', \\ u_* = (u_0,\ u_{-1},\ \ldots\ ,\ u_{-q+1})', \\ W = (W_1,\ \ldots\ ,\ W_n)'. \end{array} \tag{2.1}$$

Pour φ, θ et $\bar{\mu}$ fixés, on peut calculer

$$u_t \equiv u_t\left(\varphi,\ \theta,\ \bar{\mu} \mid W_*,\ u_*,\ W\right)\ ,\ t = 1,\ \ldots\ ,\ n.$$

La densité conjointe de $u_1,\ \ldots\ ,\ u_m$ s'écrit

$$p\left(u_1,\ \ldots\ ,\ u_n\right) = \frac{1}{\left(2\pi\right)^{n/2}\sigma^n} \exp\left\{-\frac{1}{2\sigma^2}\sum_{t=1}^{n} u_t^2\right\}\ .$$

La densité conjointe de $W_1,\ \ldots\ ,\ W_n$ étant donné W_* et u_* est :

$$\begin{aligned} L_*\left(\varphi,\ \theta,\ \bar{\mu},\ \sigma^2\right) &= p\left(W_1,\ \ldots\ ,\ W_n \mid W_*,\ u_*\right) \\ &= \frac{1}{\left(2\pi\right)^{n/2}\sigma^n} \exp\left\{-\frac{1}{2\sigma^2}\sum_{t=1}^{n} u_t\left(\varphi,\ \theta,\ \bar{\mu} \mid\ W_*,\ u_*,\ W\right)^2\right\}\ . \end{aligned}$$

La méthode de la vraisemblance maximale suggère de maximiser L_* par rapport à $\varphi,\ \varphi,\ \bar{\mu}$ et σ^2 ce qui est équivalent à maximiser

$$\ell_* \equiv \log(L_*)$$

où

$$\ell_*\left(\varphi,\ \theta,\ \bar{\mu},\ \sigma^2\right) = -\frac{n}{2}\ell n\left(2\pi\right) - \frac{n}{2}\log\left(\sigma^2\right) - \frac{S_*\left(\varphi,\ \theta,\ \bar{\mu}\right)}{2\sigma^2}$$

et

$$S_*\left(\varphi,\ \theta,\ \bar{\mu}\right) = \sum_{t=1}^{n} u_t\left(\varphi,\ \theta,\ \bar{\mu} \mid W_*,\ u_*,\ W\right)^2 .$$

Quelque soit la valeur de σ^2,

$$\text{maximiser } \ell_* \text{ est équivalent à minimiser } S_*\left(\varphi,\ \theta,\ \bar{\mu}\right).$$

Une fois $\hat{\varphi}$, $\hat{\theta}$, $\overset{\wedge}{\bar{\mu}}$ trouvés (ces derniers ne dépendent pas de σ^2), la valeur de σ^2 peut être trouvée en maximisant ℓ^* par rapport à σ^2. La condition de premier donne alors :

$$-\frac{n}{2}\frac{1}{\hat{\sigma}^2}+\frac{S_*\big(\hat{\varphi},\ \hat{\theta},\ \overset{\wedge}{\bar{\mu}}\big)}{2\hat{\sigma}^4}=0$$

d'où

$$\hat{\sigma}^2=\frac{1}{n}S_*\big(\hat{\varphi},\ \hat{\theta},\ \overset{\wedge}{\bar{\mu}}\big).$$

Le problème se trouve réduit à minimiser $S_*\left(\varphi,\ \theta,\ \bar{\mu}\right)$.

Difficulté : W_*, u_* inconnus.

La solution habituelle consiste à :

1. prendre $W_1,\ \ldots,\ W_p$ comme valeurs initiales (réduction du nombre d'observations de n à $n-p$),
$$W_*=\tilde{W}\equiv\left(W_1,\ \ldots,\ W_p\right)'; \tag{2.2}$$
2. remplacer u_* par $E\left(u_*\right)=0$.

On minimise

$$\bar{S}_*=\sum_{t=p+1}^{n}u_t\left(\varphi,\ \theta,\ \bar{\mu}\mid W_*=\tilde{W},\ u_*=0\right)^2.$$

Il s'agit là d'un problème de minimisation non linéaire.

3. Fonction de vraisemblance non-conditionnelle

On peut montrer que le logarithme de la fonction de vraisemblance non-conditionnelle de $W_1,\ \ldots,\ W_n$ a la forme :

$$\ell\left(\varphi,\ \theta,\ \bar{\mu},\ \sigma^2\right)=f\left(\varphi,\ \theta\right)-\frac{n}{2}\log\left(\sigma^2\right)-\frac{S\left(\varphi,\ \theta,\ \bar{\mu}\right)}{2\sigma^2} \tag{3.1}$$

où

$$\begin{aligned}S\left(\varphi,\ \theta,\ \bar{\mu}\right) &= \sum_{t=-\infty}^{n}\left[u_t\mid\varphi,\ \theta,\ \bar{\mu},\ W\right]^2,\\ \left[u_t\mid\varphi,\ \theta,\ \bar{\mu},\ W\right] &= E\left(u_t\mid\varphi,\ \theta,\ \bar{\mu},\ W\right)\end{aligned}$$

$$\equiv \; [u_t] \; .$$

Pour n grand ou modérément grand, $f(\varphi, \theta)$ négligeable.

Le problème se réduit alors à calculer et minimiser S.

Pour ce faire, on utilise la technique de la *prévision rétrospective (backforecasting)*.

On peut montrer que tout processus $ARMA(p, q)$ stationnaire

$$\varphi(B) W_t = \theta(B) u_t + \bar{\mu} \; , \;\; u_t \sim BB\left(0, \sigma^2\right)$$

peut aussi s'écrire

$$\varphi(F) W_t = \varphi(F) e_t + \bar{\mu} \; , \;\; e_t \sim BB\left(0, \sigma^2\right)$$

où $F = B^{-1}$ et e_t est non corrélé avec $W_{t+1}, W_{t+2}, \ldots$. On a donc

$$W_t = \varphi_1 W_{t+1} + \cdots + \varphi_p W_{t+p} + e_t - \theta_1 e_{t+1} - \cdots - \theta_q e_{t+q} + \bar{\mu}$$

d'où

$$\begin{aligned} e_t &= W_t - \varphi_1 W_{t+1} - \cdots - \varphi_p W_{t+p} \\ &\quad + \theta_1 e_{t+1} + \cdots + \theta_q e_{t+q} - \bar{\mu} \\ [e_t] &= [W_t] - \varphi_1 [W_{t+1}] - \cdots - \varphi_p [W_{t+p}] \\ &\quad + \theta_1 [e_{t+1}] + \cdots + \theta_q [e_{t+q}] - \bar{\mu}. \end{aligned}$$

On peut voir aisément que

$$[W_t] = W_t \; , \;\; t = 1, \ldots , n.$$

Si on utilise l'approximation

$$[e_t] = 0 \; , \;\; t \geq n - p + 1$$

on peut calculer

$$[e_t] \;\; , \;\; t = n - p \, , \, n - p - 1 \, , \, \ldots \, , \, 2, \, 1.$$

En particulier,

$$[e_1] = W_1 - \varphi_1 W_2 - \cdots - \varphi_p W_{p+1} + \theta_1 [e_2] + \cdots + \theta_q [e_{q+1}] - \bar{\mu}.$$

De plus

$$[e_t] = 0 \; , \;\; t \leq 0 \quad (e_t \text{ non corrélé avec } W).$$

De là, on peut calculer

$$[W_t] = \varphi_1 [W_{t+1}] + \cdots + \varphi_p [W_{t+p}]$$

$$+ [e_t] - \theta_1 [e_{t+1}] - \cdots - \theta_q [e_{t+q}] + \bar{\mu} , \;\; t = 0, -1, -2, \ldots \tag{3.2}$$

On a donc la suite $[W_t]$, $t = \ldots , -2, -1, 0, 1, \ldots , n.$

Étant donné la série

$$[W_t] \;,\; t \leq n \,,$$

on revient au modèle standard :

$$\begin{aligned} W_t &= \varphi_1 W_{t-1} + \cdots + \varphi_p W_{t-p} \\ &\quad + u_t - \theta_1 u_{t-1} - \cdots - \theta_q u_{t-q} + \bar{\mu} \,, \end{aligned} \tag{3.3}$$

$$\begin{aligned} u_t &= W_t - \varphi_1 W_{t-1} - \cdots - \varphi_p W_{t-p} \\ &\quad + \theta_1 u_{t-1} + \cdots + \theta_q u_{t-q} - \bar{\mu} \,, \end{aligned} \tag{3.4}$$

$$\begin{aligned} [u_t] &= [W_t] - \varphi_1 [W_{t-1}] - \cdots - \varphi_p [W_{t-p}] \\ &\quad + \theta_1 [u_{t-1}] + \cdots + \theta_q [u_{t-q}] - \bar{\mu}. \end{aligned} \tag{3.5}$$

Pour un processus stationnaire, on peut généralement supposer que

$$[u_t] = 0 \;,\; t < Q' \leq 0$$

ce qui permet de calculer

$$[u_t] \,,\;\; t = Q',\, Q' + 1, \ldots , -1, 0, 1, 2, \ldots , n$$

et

$$S(\varphi,\, \theta,\, \bar{\mu}) \simeq \sum_{t=Q'}^{n} [u_t]^2 \;.$$

Q' est choisi de façon à ce que $[W_t] - \mu \simeq 0.$

Ces calculs peuvent sembler complexes mais ils sont relativement faciles à programmer.

On pourrait aussi calculer le MV exact. Plus difficile à programmer ; voir Newbold (1974), Ansley (1979), Brockwell and Davis (1991, Chapter 8).

L'estimateur du maximum de vraisemblance de σ^2 est

$$\hat{\sigma}^2 = S\big(\hat{\varphi},\, \hat{\theta},\, \hat{\bar{\mu}}\big)/n.$$

Mais il peut être plus naturel d'utiliser

$$\begin{aligned}
\tilde{\sigma}^2 &= \frac{1}{n-\ell} S\big(\hat{\varphi},\, \hat{\theta},\, \overset{\wedge}{\bar{\mu}}\big) = \frac{1}{n-\ell} \sum_{t=Q'}^{n} \hat{u}_t^2 \ , \\
\ell &= \text{nombre de paramètres (à part } \sigma^2\text{) estimés.}
\end{aligned}$$

4. Tests et intervalles de confiance

Les estimateurs du maximum de vraisemblance sont asymptotiquement normaux, ce qui permet de construire des intervalles de confiance asymptotiques à partir des écart-types estimés des coefficients.

On peut tester des hypothèse du type

$$H_0 : \psi(\varphi,\, \theta) = 0\,, \tag{4.1}$$

où ψ est une fonction vectorielle de dimension r, aussi aisément en utilisant le critère du quotient de vraisemblance. Si

$$\begin{gathered}
\ell\big(\hat{\varphi},\, \hat{\theta},\, \overset{\wedge}{\bar{\mu}},\, \hat{\sigma}^2\big) \simeq -\frac{n}{2} \log\left(\hat{\sigma}^2\right) - \frac{n}{2} \\
2\left[\ell_{NC} - \ell_C\right] \underset{n\to\infty}{\longrightarrow} \chi^2(r) \ \text{ sous } H_0 \\
\ell_C = \log L \text{ contraint } (r \text{ contraintes pour } H_0) \\
\ell_{NC} = \log L \text{ non-contraint}
\end{gathered}$$

le critère du quotient de vraisemblance est donné par :

$$LR = 2\left[\ell_{NC} - \ell_C\right] = 2\left[\frac{n}{2}\log\left(\hat{\sigma}_C^2\right) - \frac{n}{2}\log\left(\hat{\sigma}_{NC}^2\right)\right] = n \log\left(\hat{\sigma}_C^2 / \hat{\sigma}_{NC}^2\right).$$

On rejette H_0 lorsque $LR \geq \chi^2(\alpha;\, r)$où $P[\chi^2(r) \geq \chi^2(\alpha;\, r)] = \alpha$.

5. Notes bibliographiques

Plusieurs ont discuté l'estimation des modèles ARMA ; voir notamment Box and Jenkins (1976, Chapter 7) et Brockwell and Davis (1991, Chapter 8).

Références

ANSLEY, C. F. (1979) : "An Algorithm for the Exact Likelihood of a Mixed Autoregressive-Moving Average Process," *Biometrika*, 66, 59–66.

BOX, G. E. P., AND G. M. JENKINS (1976) : *Time Series Analysis : Forecasting and Control*. Holden-Day, San Francisco, second edn.

BROCKWELL, P. J., AND R. A. DAVIS (1991) : *Time Series : Theory and Methods*. Springer-Verlag, New York, second edn.

NEWBOLD, P. (1974) : "The Exact Likelihood Function for a Mixed Autoregressive-Moving Average Process," *Biometrika*, 61, 423–426.