Estimation of ARMA models by maximum likelihood *

Jean-Marie Dufour †
McGill University

First version: February 1981
Revised: February 1991, September 2000
This version: February 11, 2008
Compiled: February 11, 2008, 3:00pm

* This work was supported by the William Dow Chair in Political Economy (McGill University), the Canada Research Chair Program (Chair in Econometrics, Université de Montréal), the Bank of Canada (Research Fellowship), a Guggenheim Fellowship, a Konrad-Adenauer Fellowship (Alexander-von-Humboldt Foundation, Germany), the Institut de finance mathématique de Montréal (IFM2), the Canadian Network of Centres of Excellence [program on *Mathematics of Information Technology and Complex Systems* (MITACS)], the Natural Sciences and Engineering Research Council of Canada, the Social Sciences and Humanities Research Council of Canada, the Fonds de recherche sur la société et la culture (Québec), and the Fonds de recherche sur la nature et les technologies (Québec), and a Killam Fellowship (Canada Council for the Arts).2

† William Dow Professor of Economics, McGill University, Centre interuniversitaire de recherche en analyse des organisations (CIRANO), and Centre interuniversitaire de recherche en économie quantitative (CIREQ). Mailing address: Department of Economics, McGill University, Leacock Building, Room 519, 855 Sherbrooke Street West, Montréal, Québec H3A 2T7, Canada. TEL: (1) 514 398 8879; FAX: (1) 514 398 4938; e-mail: jean-marie.dufour@mcgill.ca . Web page: http://www.jeanmariedufour.com

Contents

1. Model

Let

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

X_t follows the model :

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

where

$$\nabla = 1 - B, \tag{1.3}$$

$$\varphi\left(B\right) = 1 - \varphi_1 B - \cdots - \varphi_p B^p\,,\ \varphi_p \neq 0, \tag{1.4}$$

$$\theta\left(B\right) = 1 - \theta_1 B - \cdots - \theta_q B^q\,,\ \theta_q \neq 0, \tag{1.5}$$

$$\varphi\left(B\right) \text{ and } \theta\left(B\right) \text{ have no common root,} \tag{1.6}$$

$$u_t \sim BB\left(0,\, \sigma^2\right)\ . \tag{1.7}$$

Series of $N = n + d$ values : $X_{-d+1}, \ldots, X_0, X_1, \ldots, X_n$ $p + q + 2$ parameters to estimate.

Set :

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

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

where X_t has possibly been transformed through a power or logarithmic transformation. We can write:

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

If we set $\tilde{W}_t = W_t - \mu$, where

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

Suppose that $u_t \overset{i.i.d}{\sim} N\left(0,\, \sigma^2\right).$

Difficulty: at $t = 1,\, W_{t-1},\, \ldots\,,\, W_{t-p},\, u_{t-1},\, \ldots\,,\, u_{t-q}$ are unknown.

We can use here two main approaches:

1. maximize the conditional likelihood function;

2. maximize the unconditional .

2. Conditional likelihood function

Let

$$\begin{array}{l} W_* : p \text{ observations of } W_t \text{ before series beginning,} \\ u_* : q \text{ noises } u_t, \\ W_* = \left(W_0,\, W_{-1},\, \ldots\,,\, W_{-p+1}\right)', \\ u_* = \left(u_0,\, u_{-1},\, \ldots\,,\, u_{-q+1}\right)', \\ W = \left(W_1,\, \ldots\,,\, W_n\right)'. \end{array} \tag{2.1}$$

For φ, θ and $\bar{\mu}$ given, we can compute:

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

The joint density of $u_1,\, \ldots\,,\, u_m$ can be written

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

The joint density of $W_1,\, \ldots\,,\, W_n$ given W_* and u_* is:

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

The maximum likelihood method suggests that we maximize $L_*\, with$respect to $\varphi,\, \varphi,\, \bar{\mu}\, and\, \sigma^2$, which is equivalent to maximizing

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

where

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

and

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

Irrespective of the value of σ^2,

maximizing ℓ_* is equivalent to minimizing $S_*(\varphi,\,\theta,\,\bar{\mu})\,.$

Once $\hat{\varphi},\,\hat{\theta},\,\hat{\bar{\mu}}$ have been estimated (these do not depend on σ^2), the value of σ^2 can be obtained by maximizing ℓ^* with respect to σ^2. The first-order condition then yields:

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

hence

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

The problem is then reduced to minimizing $S_*(\varphi,\,\theta,\,\bar{\mu})\,.$

Difficulty : $W_*,\,u_*$ unknown.

The usual solution consists in:

1. taking $W_1,\,\ldots\,,\,W_p$ as initial values (this reduces the number of observations from n to $n-p$),
$$W_* = \tilde{W} \equiv (W_1,\,\ldots\,,\,W_p)'\,; \tag{2.2}$$
2. replacing u_* by $E(u_*) = 0$.

We minimize

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

This is a nonlinear minimization problem.

3. Unconditional likelihood function

One can show that the logarithm of the unconditional likelihood function of $W_1,\,\ldots\,,\,W_n$ has the form:

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

where

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

For n large (or moderately large), $f(\varphi, \theta)$ is negligible.

The problem is then reduced to computing and minimizing S.

To do this, we use the technique of *backforecasting*.

One can show that every stationary $ARMA(p, q)$ process

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

can also be written

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

where $F = B^{-1}$ and e_t is uncorrelated with $W_{t+1}, W_{t+2}, \ldots$. We then have

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

where

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

We can see easily that

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

If we use the approximation

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

we can compute

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

In particular,

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

Furthermore

$$[e_t] = 0\ ,\ t \leq 0 \quad (e_t \text{ uncorrelated with } W).$$

From there, we can compute

$$[W_t] = \varphi_1 [W_{t+1}] + \cdots + \varphi_p [W_{t+p}]$$
$$+ [e_t] - \theta_1 [e_{t+1}] - \cdots - \theta_q [e_{t+q}] + \bar{\mu}\,,\ \ t = 0,\, -1,\, -2,\, \ldots \tag{3.2}$$

We then get the sequence $[W_t]\,,\, t = \ldots\,,\, -2,\, -1,\, 0,\, 1,\, \ldots\,,\, n.$

Given the series

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

we go back to the standard model:

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

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

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

For a stationary process, we can suppose that

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

which allows one to compute

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

and

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

Q' is chosen so that $[W_t] - \mu \simeq 0$.

These calculations may appear complex, but they are relatively easy to program.

One could also compute the exact ML estimator. But this is more difficult to program; see Newbold (1974), Ansley (1979), Brockwell and Davis (1991, Chapter 8).

The maximum likelihood estimator of σ^2 is

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

But it is more natural to use

$$\begin{aligned} \tilde{\sigma}^2 &= \frac{1}{n-\ell} S\big(\hat{\varphi},\, \hat{\theta},\, \overset{\wedge}{\bar{\mu}}\big) = \frac{1}{n-\ell} \sum_{t=Q'}^{n} \hat{u}_t^2 \,, \\ \ell &= \text{number of parameters (except for } \sigma^2\text{) estimated.} \end{aligned}$$

4. Tests and confidence intervals

The maximum likelihood estimators are asymptotically normal, which allows one to build asymptotic confidence intervals based on estimated coefficient standard errors.

We can test hypotheses of the type

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

where ψ is a vector function of dimension r, relatively easily by using the likelihood ratio criterion. If

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

the likelihood ratio is given by

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

We reject H_0 when $LR \geq \chi^2\left(\alpha;\, r\right)$ where $P[\chi^2\left(r\right) \geq \chi^2\left(\alpha;\, r\right)] = \alpha$.

5. Bibliographic notes

Several textbooks discuss the estimation of ARMA models; see, for example Box and Jenkins (1976, Chapter 7) et Brockwell and Davis (1991, Chapter 8).

References

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.