Foreasting of stationary and ARIMA processes *

Jean-Marie Dufour †
McGill University

First version: March 1985
Revised: February 2008
This version: February 19, 2008
Compiled: February 19, 2008, 8:20pm

* 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).

† 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. Wiener-Kolmogorov formula

According to Wold's decomposition theorem, a second-order stationary process (with mean zero) can be written in the following form:

$$X_t = Y_t + D_t$$

where

$$\begin{aligned} Y_t &= \sum_{j=0}^{\infty} d_j u_{t-j}\,, \quad d_0 = 1,\ \{u_t\} \sim BB\left(0,\, \sigma^2\right)\,, \\ &D_t \text{ is deterministic}\,, \\ u_t &= X_t - P_L\left(X_t \mid X_{t-1},\, X_{t-2},\, \ldots\right)\,. \end{aligned}$$

Consequently,

$$P_L\left(u_t \mid X_{t-1},\, X_{t-2},\, \ldots\right) = 0$$

or, more generally,

$$P_L\left(u_t \mid X_{t-\ell},\, X_{t-\ell-1},\, \ldots\right) = 0 \quad , \quad \forall \ell \geq 1.$$

If X_t is a strictly indeterministic process,

$$X_t = \sum_{j=0}^{\infty} d_j u_{t-j}\,,$$

we have:

$$\begin{aligned} P_L\left(X_t \mid X_{t-1},\, X_{t-2},\, \ldots\right) &= P_L\left(u_t \mid X_{t-1},\, X_{t-2},\, \ldots\right) \\ &+ P_L\left[\sum_{j=1}^{\infty} d_j u_{t-j} \mid X_{t-1},\, X_{t-2},\, \ldots\right] \\ &= \sum_{j=1}^{\infty} d_j u_{t-j}\,. \end{aligned}$$

If we furthermore suppose that the $\{u_t\}$ are independent,

$$\begin{aligned} \mathsf{E}\left(u_t \mid X_{t-1},\, X_{t-2},\, \ldots\right) &= \mathsf{E}\left(u_t \mid u_{t-1},\, u_{t-2},\, \ldots\right) = 0\,, \\ \mathsf{E}\left(X_t \mid X_{t-1},\, X_{t-2},\, \ldots\right) &= \sum_{j=1}^{\infty} d_j u_{t-j} \end{aligned}$$

$$= \quad P_L\left(X_t \mid X_{t-1},\, X_{t-2},\, \ldots\right)\ .$$

We have the best prediction in the mean-square-error (MSE) sense.

Let $\{X_t\}$ a weakly stationary indeterministic process:

$$X_t = \sum_{j=0}^{\infty} d_j u_{t-j} = d\left(B\right) u_t, \tag{1.1}$$

where

$$d_0 = 1\ , \quad d\left(B\right) = \sum_{j=0}^{\infty} d_j B^j\ . \tag{1.2}$$

Let us denote:

$$P_{t-j}X_t = P\left(X_t \mid X_{t-j},\, X_{t-j-1},\, \ldots\right)\ .$$

Then,

$$\begin{aligned}
P_{t-j}X_t &= X_t \ \text{ for } \ j \leq 0\ , \\
u_t &= X_t - P\left(X_t \mid X_{t-1},\, X_{t-2},\, \ldots\right) \\
&= X_t - P_{t-1}X_t \\
P_{t-1}X_t &= \sum_{j=0}^{\infty} d_j P_{t-1} u_{t-j} = \sum_{j=1}^{\infty} d_j u_{t-j} \\
&= \left(\frac{d\left(B\right)}{B}\right)_+ u_{t-1}
\end{aligned}$$

where we define

$$\left(\sum_{j=-\infty}^{+\infty} h_j B^j\right)_+ = \sum_{j=0}^{\infty} h_j B^j\ ,$$

$$\begin{aligned}
\left[\frac{d\left(B\right)}{B}\right]_+ &= \left(d_0 B^{-1} + d_1 B^0 + d_2 B^1 + \cdots\right) \\
&= \left(\sum_{j=0}^{\infty} d_j B^{j-1}\right)_+ \\
&= \sum_{j=1}^{\infty} d_j B^{j-1}\ .
\end{aligned} \tag{1.3}$$

Similarly, we get at lag ℓ,

$$\begin{aligned} P_{t-\ell}X_t &= \sum_{j=0}^{\infty} d_j P_{t-\ell} u_{t-j} = \sum_{j=\ell}^{\infty} d_j u_{t-j} \\ &= \left(\frac{d\,(B)}{B^{\ell}} \right)_+ u_{t-\ell} \end{aligned}$$

or, equivalently,

$$P_t X_{t+\ell} = \left(\frac{d\,(B)}{B^{\ell}} \right)_+ u_t = \sum_{j=\ell}^{\infty} d_j u_{t+\ell-j}\ . \tag{1.4}$$

The formula (1.4) is called the Wiener-Kolmogorov formula for linear prediction ℓ periods ahead. We see that $P_t X_{t+\ell}$ can be computed by dropping the most recent l terms from the Wold decomposition:

$$\begin{aligned} X_t &= \sum_{j=0}^{\infty} d_j u_{t-j} \\ &= \sum_{j=0}^{l-1} d_j u_{t+\ell-j} + \sum_{j=\ell}^{\infty} d_j u_{t+\ell-j} \\ &= e_t(l) + P_t X_{t+\ell}\ . \end{aligned} \tag{1.5}$$

If X_t is invertible, we can write

$$u_t = \frac{1}{d\,(B)} X_t$$

hence

$$\begin{aligned} P_{t-\ell}X_t &= \left(\frac{d\,(B)}{B^{\ell}} \right)_+ \frac{1}{d\,(B)} X_{t-\ell}\ , \\ P_t X_{t+\ell} &= \left(\frac{d\,(B)}{B^{\ell}} \right)_+ \frac{1}{d\,(B)} X_t\ ,\ \ell \geq 1\ . \end{aligned}$$

1.1 Example For an AR(1) process,

$$X_t = \varphi_1 X_{t-1} + u_t\,, \quad |\varphi_1| < 1 \tag{1.6}$$

we have the representation

$$X_t = \frac{1}{1-\varphi_1 B} u_t\,, \tag{1.7}$$

hence

$$\begin{aligned} P_{t-1}X_t &= \varphi_1 P_{t-1}X_{t-1} + P_{t-1}u_t \\ &= \varphi_1 X_{t-1}\,, \end{aligned} \tag{1.8}$$

$$\begin{aligned} P_{t-\ell}X_t &= \varphi_1 P_{t-\ell}X_{t-1} \\ &= \varphi_1 P_{t-\ell}\left(\varphi_1 X_{t-2} + u_{t-1}\right) \\ &= \varphi_1^2 P_{t-\ell}X_{t-2} \\ &= \varphi_1^{\ell} X_{t-\ell}\,. \end{aligned}$$

If we use the Wiener-Kolmogorov formula, we get:

$$\begin{aligned} P_{t-\ell}X_t &= \left[B^{-\ell}\frac{1}{1-\varphi_1 B}\right]_+ u_{t-\ell} \\ &= \left[B^{-\ell}\left(1+\varphi_1 B + \varphi_1^2 B^2 + \cdots\right)\right]_+ \left(1-\varphi_1 B\right) X_{t-\ell} \\ &= \varphi_1^{\ell}\left(1+\varphi_1 B + \varphi_1^2 B^2 + \cdots\right)\left(1-\varphi_1 B\right) X_{t-\ell} \\ &= \frac{\varphi_1^{\ell}}{1-\varphi_1 B}\left(1-\varphi_1 B\right) X_{t-\ell} \\ &= \varphi_1^{\ell} X_{t-\ell}\,. \end{aligned}$$

1.2 Example For an MA(1)process,

$$X_t = \left(1-\theta_1 B\right) u_t\,, \quad |\theta_1| < 1, \tag{1.9}$$

we have

$$u_t = \frac{1}{1-\theta_1 B} X_t\,, \tag{1.10}$$

hence the following forecasts: for $\ell = 1$,

$$\begin{aligned} P_{t-1}X_t &= \left[B^{-1}\left(1-\theta_1 B\right)\right]_+ \frac{1}{1-\theta_1 B} X_{t-1} \\ &= \frac{-\theta_1}{1-\theta_1 B} X_{t-1} \\ &= -\theta_1 \sum_{i=0}^{\infty} \theta_1^i X_{t-1-i} \\ &= -\sum_{i=1}^{\infty} \theta_1^i X_{t-i}\,; \end{aligned} \tag{1.11}$$

pour $\ell \geq 2$

$$P_{t-\ell}X_t = \left[B^{-\ell}\left(1-\theta_1 B\right)\right]_+ \frac{1}{1-\theta_1 B}X_{t-\ell} = 0 \; .$$

1.3 Example For an ARMA$(1,1)$ process of the form

$$\begin{aligned} \left(1-\varphi_1 B\right)X_t &= \left(1-\theta_1 B\right)u_t \; , \\ \mid \varphi_1 \mid < 1 \; , \; \mid \theta_1 \mid &< 1 \, , \end{aligned} \tag{1.12}$$

the Wold representation can be written:

$$\begin{aligned} X_t &= \frac{\left(1-\theta_1 B\right)}{\left(1-\varphi_1 B\right)}u_t \\ &= d\left(B\right)u_t \, . \end{aligned} \tag{1.13}$$

From the latter, we then get the following forecasts: at the horizon 1,

$$P_{t-1}X_t = \left[\frac{d\left(B\right)}{B}\right]_+ u_{t-1} = \left[\frac{d\left(B\right)}{B}\right]_+ \frac{1}{d\left(B\right)}X_{t-1} \tag{1.14}$$

where

$$\begin{aligned} \left[\frac{d\left(B\right)}{B}\right]_+ &= \left[B^{-1}\frac{\left(1-\theta_1 B\right)}{\left(1-\varphi_1 B\right)}\right]_+ \\ &= \left[\frac{B^{-1}}{1-\varphi_1 B} - \frac{\theta_1}{1-\varphi_1 B}\right]_+ \\ &= \left[B^{-1}\left[1+\varphi_1 B\left(1+\varphi_1 B+\varphi_1^2 B^2+\cdots\right)\right] - \frac{\theta_1}{1-\varphi_1 B}\right]_+ \\ &= \left[B^{-1}+\frac{\varphi_1}{1-\varphi_1 B} - \frac{\theta_1}{1-\varphi_1 B}\right]_+ = \frac{\left(\varphi_1-\theta_1\right)}{1-\varphi_1 B}, \end{aligned} \tag{1.15}$$

so that

$$\begin{aligned} P_{t-1}X_t &= \frac{\left(\varphi_1-\theta_1\right)}{1-\varphi_1 B}\left(\frac{1-\varphi_1 B}{1-\theta_1 B}\right)X_{t-1} \\ &= \frac{\left(\varphi_1-\theta_1\right)}{1-\theta_1 B}X_{t-1}; \end{aligned} \tag{1.16}$$

at lag ℓ,

$$P_{t-\ell}X_t = \left[\frac{d\left(B\right)}{B^\ell}\right]_+ u_{t-\ell} = \left[\frac{d\left(B\right)}{B^\ell}\right]_+ \frac{1}{d\left(B\right)}X_{t-\ell} \tag{1.17}$$

where

$$\begin{aligned} \left[\frac{d(B)}{B^{\ell}}\right]_{+} &= \left[B^{-\ell}\frac{(1-\theta_1 B)}{(1-\varphi_1 B)}\right]_{+} \\ &= \left[\frac{B^{-\ell}}{1-\varphi_1 B} - \frac{\theta_1 B^{-(\ell-1)}}{(1-\varphi_1 B)}\right] \\ &= \left[\frac{\varphi_1^{\ell}}{1-\varphi_1 B} - \frac{\theta_1 \varphi_1^{\ell-1}}{1-\varphi_1 B}\right] \\ &= \frac{\varphi_1^{\ell-1}(\varphi_1-\theta_1)}{1-\varphi_1 B}, \end{aligned} \tag{1.18}$$

hence

$$\begin{aligned} P_{t-\ell}X_t &= \frac{(\varphi_1-\theta_1)\,\varphi_1^{\ell-1}}{1-\varphi_1 B}\left(\frac{1-\varphi_1 B}{1-\theta_1 B}\right) X_{t-\ell} \\ &= \frac{\varphi_1^{\ell-1}(\varphi_1-\theta_1)}{1-\theta_1 B} X_{t-\ell} \end{aligned} \tag{1.19}$$

or

$$P_t X_{t+\ell} = \frac{\varphi_1^{\ell-1}(\varphi_1-\theta_1)}{1-\theta_1 B} X_t \ .$$

2. Chain rule for prediction

Let

$$P_t X_{t+1} = \sum_{j=0}^{\infty} h_j X_{t-j} \ .$$

Then,

$$P_{t+\ell}X_{t+\ell+1} = \sum_{j=0}^{\infty} h_j X_{t+\ell-j}$$
$$= h_0 X_{t+\ell} + h_1 X_{t+\ell-1} + \cdots + h_\ell X_t + h_{\ell+1} X_{t-1} + \cdots$$

$$P_t\left[P_{t+\ell}X_{t+\ell+1}\right] = P_t X_{t+\ell+1}$$
$$= h_0 P_t X_{t+\ell} + h_1 P_t X_{t+\ell-1} + \cdots + h_{\ell-1} P_t X_{t+1}$$
$$+ h_\ell X_t + h_{\ell+1} X_{t-1} + \cdots$$

$$= \sum_{i=0}^{\ell-1} h_i P_t X_{t+\ell-i} + \sum_{i=\ell}^{\infty} h_i X_{t+\ell-i}$$
$$= \sum_{i=0}^{\ell-1} h_i P_t X_{t+\ell-i} + \sum_{i=0}^{\infty} h_{i+\ell} X_{t-i} \; .$$

This formula allows one to compute predictions several steps ahead from one-step ahead predictions.

3. Properties of prediction errors

Let

$$P_t X_{t+\ell} = P\left(X_{t+\ell} \mid X_t,\, X_{t-1},\, \ldots\right) \; .$$

If

$$X_t = \sum_{j=0}^{\infty} d_j u_{t-j} \; , \tag{3.1}$$

where

$$\begin{aligned} d_0 &= 1, & (3.2)\\ u_t &= X_t - P_{t-1} X_t \, , & (3.3)\\ \mathsf{V}\left(u_t\right) &= \sigma^2, & (3.4) \end{aligned}$$

then

$$\begin{aligned} X_{t+\ell} &= \sum_{j=0}^{\infty} d_j u_{t+\ell-j} \, , \\ P_t X_{t+\ell} &= \sum_{j=\ell}^{\infty} d_j u_{t+\ell-j} \; . \end{aligned}$$

$$\begin{aligned} e_t\left(\ell\right) &\equiv X_{t+\ell} - P_t X_{t+\ell} \\ &= \sum_{j=0}^{\ell-1} d_j u_{t+\ell-j} \\ &= u_{t+\ell} + d_1 u_{t+\ell-1} + \cdots + d_{\ell-1} u_{t+1} \end{aligned}$$

follows a $MA\left(\ell-1\right)$ process, hence

$$\begin{aligned}\mathsf{V}\left[e_t\left(\ell\right)\right] &= \sigma^2\left[1+d_1^2+\cdots+d_{\ell-1}^2\right]\\ &= \sigma^2\sum_{i=0}^{\ell-1}d_i^2\,,\end{aligned}$$

where $d_0=1$. Consequently,

$$e_t\left(1\right)=u_{t+1}\,. \tag{3.5}$$

One-step ahead prediction errors associated with optimal are uncorrelated between each others. Further,

$$\begin{aligned}\mathsf{E}\left[e_t\left(\ell\right)e_{t+k}\left(\ell\right)\right] &= \sigma^2\sum_{i=0}^{\ell-k-1}d_id_{i+k}\\ &= \sigma^2\sum_{i=k}^{\ell-1}d_id_{i-k}\,,\quad 0\leq k\leq\ell-1,\end{aligned} \tag{3.6}$$

$$\mathsf{E}\left[e_t\left(\ell\right)e_{t+k}\left(\ell\right)\right]=\begin{cases}\sigma^2\sum_{i=k}^{\ell-1}d_id_{i-k}\,, & \text{if } 0\leq k\leq\ell-1\,,\\ 0\,, & \text{if } k\geq\ell\,,\end{cases} \tag{3.7}$$

$$\mathsf{Corr}\left[e_t\left(\ell\right),\,e_{t+k}\left(\ell\right)\right]=\begin{cases}\frac{\sum_{i=k}^{\ell-1}d_id_{i-k}}{\sum_{j=0}^{\ell-1}d_j^2}\,, & \text{if } 0\leq k\leq\ell-1\,,\\ 0\,, & \text{if } k\geq\ell\,,\end{cases} \tag{3.8}$$

$$\begin{aligned}\mathsf{Cov}\left[e_t\left(\ell\right),\,e_t\left(\ell+j\right)\right] &= \mathsf{E}\left[e_t\left(\ell\right)e_t\left(\ell+j\right)\right]\\ &= \sigma^2\sum_{i=0}^{\ell-1}d_id_{i+j}\neq 0\,,\quad \text{for } j\geq 0\,.\end{aligned} \tag{3.9}$$

The covariances between prediction errors at different horizons are given by:

$$\begin{aligned}C\left[e_t\left(\ell\right),\,e_t\left(\ell+j\right)\right] &= \mathsf{E}\left[e_t\left(\ell\right)e_t\left(\ell+j\right)\right]\\ &= \sigma^2\sum_{i=0}^{\ell-1}d_id_{i+j}\,.\end{aligned}$$

4. Prediction with ARIMA models

Suppose X_t is an ARIMA process of the form:

$$\varphi_p(B)(1-B)^d X_t = \theta_q(B) u_t + \bar{\mu}, \tag{4.1}$$
$$u_t \sim BB\left(0,\ \sigma^2\right), \tag{4.2}$$

If we denote

$$\varphi(B) = \varphi_p(B)(1-B)^d, \tag{4.3}$$

we can write

$$\varphi(B) X_t = \theta_q(B) u_t + \bar{\mu}, \tag{4.4}$$

or, equivalently,

$$\begin{aligned} &\left[1 - \varphi_1 B - \cdots - \varphi_{p+d} B^{p+d}\right] X_t \\ &= \left[1 - \theta_1 B - \cdots - \theta_q B^q\right] u_t + \bar{\mu}, \end{aligned} \tag{4.5}$$

hence the *difference-equation representation* of X_t:

$$\begin{aligned} X_t &= \varphi_1 X_{t-1} + \cdots + \varphi_{p+d} X_{t-p-d} \\ &\quad + u_t - \theta_1 u_{t-1} - \cdots - \theta_q u_{t-q} + \bar{\mu} \end{aligned} \tag{4.6}$$

or

$$X_t = \sum_{i=1}^{p+d} \varphi_i X_{t-i} - \sum_{j=1}^{q} \theta_j u_{t-j} + \bar{\mu} + u_t. \tag{4.7}$$

4.1 Example For an ARIMA$(1,\ 1,\ 0)$ process, we have:

$$\begin{aligned} (1-\phi_1 B)(1-B) X_t &= u_t, \\ \left[1 - \varphi_1 B + \varphi_2 B^2\right] X_t &= \left[1 - (\phi_1 + 1) B + \phi_1 B^2\right] X_t = u_t, \\ \varphi_1 &= (\phi_1 + 1),\ \varphi_2 = -\phi_1. \end{aligned}$$

On applying the projection operator P_t on both sides of the equation (4.7), we obtain:

$$\begin{aligned} P_t X_{t+1} &= \sum_{i=1}^{p+d} \varphi_i P_t X_{t+1-i} - \sum_{j=0}^{q} \theta_j P_t u_{t+1-j} + \bar{\mu} \\ &= \sum_{i=1}^{p+d} \varphi_i X_{t+1-i} - \sum_{j=1}^{q} \theta_j u_{t+1-j} + \bar{\mu}, \end{aligned}$$

$$P_t X_{t+2} \;=\; \varphi_1 P_t X_{t+1} + \sum_{i=2}^{p+d} \varphi_i X_{t+2-i} - \sum_{j=2}^{q} \theta_j u_{t+2-j} + \bar{\mu}\,,$$

and, more generally,

$$P_t X_{t+\ell} = \sum_{i=1}^{\ell-1} \varphi_i P_t X_{t+\ell-i} + \sum_{i=\ell}^{p+d} \varphi_i X_{t+\ell-i} - \sum_{j=\ell}^{q} \theta_j u_{t+\ell-j} + \bar{\mu}\ .$$

On noting that

$$P_t u_{t+\ell} = \begin{cases} 0\,, & \text{if } \ell \geq 1\,, \\ u_{t+\ell}\ , & \text{if } \ell \leq 0\,, \end{cases}$$

we then see that

$$P_t X_{t+\ell} = \sum_{i=1}^{p+d} \varphi_i P_t X_{t+\ell-i} + P_t u_{t+\ell} - \sum_{j=1}^{q} \theta_j P_t u_{t+\ell-j} + \bar{\mu} \tag{4.1}$$

and

$$\begin{aligned} \varphi(B)\, P_t X_{t+\ell} \;&=\; \theta(B)\, P_t u_{t+\ell} + \bar{\mu}\,, \\ \varphi(B)\, X_{t+\ell} \;&=\; \theta(B)\, u_{t+\ell} + \bar{\mu}\,, \\ \varphi(B)\left(X_{t+\ell} - P_t X_{t+\ell}\right) \;&=\; \theta(B)\left(u_{t+\ell} - P_t u_{t+\ell}\right)\ . \end{aligned}$$

Let

$$\psi(B) = 1 + \psi_1 B + \psi_2 B^2 + \cdots = \sum_{j=0}^{\infty} \psi_j B^j\,,$$
$$\varphi(B)\,\psi(B) = \theta(B)\ ,$$
$$e_t(\ell) = X_{t+\ell} - P_t X_{t+\ell}\,, \quad \ell \geq 1\,,$$
$$\varphi(B)\, e_t(\ell) = \varphi(B)\,\psi(B)\left(u_{t+\ell} - P_t u_{t+\ell}\right)\ .$$

If we note that

$$\begin{aligned} e_t(\ell) \;&=\; 0\,,\ \ \ell \leq 0 \\ u_{t+\ell} - P_t u_{t+\ell} \;&=\; \begin{cases} 0\,, & \ell \leq 0 \\ u_{t+\ell}\,, & \ell \geq 1 \end{cases} \end{aligned}$$

we can simplify $\varphi(B)$ on both sides and get

$$\begin{aligned} e_t(\ell) &= \psi(B)(u_{t+\ell} - P_t u_{t+\ell}) \\ &= u_{t+\ell} + \psi_1 u_{t+\ell-1} + \cdots + \psi_{\ell-1} u_{t+1} \\ &= \sum_{i=0}^{\ell-1} \psi_i u_{t+\ell-i}\,, \end{aligned}$$

where $\psi_0 = 1$. It follows that

$$\begin{aligned} \mathsf{E}[e_t(\ell)] &= 0 \\ \mathsf{V}[e_t(\ell)] &= \mathsf{V}(\ell) = \sigma^2 \left[1 + \sum_{j=1}^{\ell-1} \psi_j^2\right] . \end{aligned}$$

One-step ahead prediction errors $e_t(1)$ are uncorrelated between each other. More generally,

$$\mathsf{E}[e_t(\ell)\, e_{t-j}(\ell)] = \begin{cases} \sigma^2 \sum_{i=j}^{\ell-1} \psi_i \psi_{i-j}\,, & \text{if } 0 \le j \le \ell - 1\,, \\ 0\,, & \text{if } \mid j \mid \ge \ell\,, \end{cases} \tag{4.8}$$

$$\mathsf{E}[e_t(\ell)\, e_t(\ell + j)] = \left(\sum_{i=0}^{\ell-1} \psi_i \psi_{i+j}\right) \sigma^2 .$$

If we assume that $u_t \overset{ind}{\sim} N[0,\, \sigma_a^2]$

$$e_t(\ell) \sim N\left[0,\, \left\{1 + \sum_{j=1}^{\ell-1} \psi_j^2\right\} \sigma^2\right] ,$$

we can compute confidence intervals for predictions:

$$\mathsf{P}\left[P_t X_{t+\ell} - c_{\alpha/2} \varDelta_\ell \le X_{t+\ell} \le P_t X_{t+\ell} + c_{\alpha/2} \varDelta_\ell\right] = 1 - \alpha$$

where

$$\varDelta_\ell^2 = \left\{1 + \sum_{j=1}^{\ell-1} \psi_j^2\right\} \sigma^2 , \tag{4.9}$$

$$\mathsf{P}[N(0,\, 1) \ge c_\alpha] = \alpha .$$

If we compute predictions at different horizons ℓ,we obtain a *prediction function*:

$$P_t X_{t+\ell} \;\equiv\; \hat{X}_t(\ell) \;,\; \ell = 1,\, 2,\, 3,\, \ldots\,,$$

$$\begin{aligned}\varphi(B) X_t &= \theta(B) u_t + \bar{\mu}\,, \\ \varphi(B) \hat{X}_t(\ell) &= \theta(B) P_t u_{t+\ell} + \bar{\mu}\,, \\ \varphi_p(B)(1-B)^d \hat{X}_t(\ell) &= \theta(B) P_t u_{t+\ell} + \bar{\mu}\,.\end{aligned}$$

If $d = 0$ and ℓ is large, we have:

$$\begin{aligned}\varphi_p(B) \hat{X}_t(\ell) &\simeq \bar{\mu}\,, \\ \hat{X}_t(\ell) &\simeq \mu \equiv \frac{\bar{\mu}}{\varphi_p(B)} = \frac{\bar{\mu}}{1-\varphi_1-\cdots-\varphi_p}\,.\end{aligned}$$

If $d = 1$ and ℓ is large,

$$\varphi_p(B)(1-B)\hat{X}_t(\ell) \simeq \bar{\mu}$$

$$(1-B)\hat{X}_t(\ell) \simeq \mu \equiv \frac{\bar{\mu}}{\varphi_p(B)}$$

$$\hat{X}_t(\ell) \simeq \mu_0 + \mu\ell \quad \text{Arithmetic progression.}$$

If $d = 2$ and ℓ is large,

$$\hat{X}_t(\ell) \simeq \mu_0 + \mu_1\ell + \mu_2\ell^2\,.$$

5. Bibliographic notes

The reader will find general discussions of prediction based on ARIMA models in Box and Jenkins (1976, Sections 5.1-5.5, 5.7), Brockwell and Davis (1991, Sections 5.1-5.5) and Hamilton (1994, Chap. 4). On prediction for stationary processes and the Wiener-Kolmogorov formula, see also Wiener (1949), Whittle (1983), Whiteman (1983) and Sargent (1987).

References

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

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

Hamilton, J. D. (1994), *Time Series Analysis*, Princeton University Press, Princeton, New Jersey.

Sargent, T. J. (1987), *Macroeconomic Theory*, second edn, Academic Press, New York .

Whiteman, C. H. (1983), *Linear Rational Expectations Models*, University of Minnesota Press, Minneapolis, MN.

Whittle, P. (1983), *Prediction and Regulation by Linear Least Square Methods*, second edn, University of Minnesota Press, New York.

Wiener, N. (1949), *Time Series*, The MIT Press, Cambridge, MA.