LECTURE 6

Forecasting with ARMA Models

If the nonstationarity of a time series can be attributed to the presence of d unit roots in the autoregressive operator, then the series can be forecast by forecasting its dth difference. With the help of d initial conditions, the forecasts of the difference can be aggregated to generate a forecast of the level of the series.

Three Forms of Forecasting Equations

An ARMA model may be represented in three different ways

$$\alpha(L)y(t) = \mu(L)\varepsilon(t), \qquad \textit{Difference Equation Form}$$

$$y(t) = \frac{\mu(L)}{\alpha(L)}\varepsilon(t) = \psi(L)\varepsilon(t), \qquad \textit{Moving-Average Form}$$

$$\frac{\alpha(L)}{\mu(L)}y(t) = \pi(L)y(t) = \varepsilon(t). \qquad \textit{Autoregressive Form}$$

Here

$$\psi(L) = \{1 + \psi_1 L + \psi_2 L^2 + \cdots\} \qquad \text{and}$$

$$\pi(L) = \{1 - \pi_1 L - \pi_2 L^2 - \cdots\}$$

stand for the series expansions of the respective rational operators.

In developing the theory of forecasting, we may consider an infinite information set $\mathcal{I}_t = \{y_t, y_{t-1}, y_{t-2}, \ldots\}$. Knowing the parameters in $\alpha(L)$ and $\mu(L)$ enables us to recover the sequence $\{\varepsilon_t, \varepsilon_{t-1}, \varepsilon_{t-2}, \ldots\}$ from the sequence $\{y_t, y_{t-1}, y_{t-2}, \ldots\}$ and vice versa; so either of these constitute the information set. This equivalence implies that the forecasts may be expressed in terms $\{y_t\}$ or in terms $\{\varepsilon_t\}$ or as a combination of the elements of both sets.

The one-step-ahead forecasts corresponding to these forms are given by

$$\begin{aligned}\hat{y}_{t+1|t} = &-\{\alpha_1 y_t + \alpha_2 y_{t-1} + \cdots + \alpha_p y_{t-p+1}\} \quad \textit{Difference Equation Form}\\ &+ \{\mu_1\varepsilon_t + \mu_2\varepsilon_{t-1} + \cdots + \mu_q\varepsilon_{t-q+1}\},\end{aligned}$$

$$\hat{y}_{t+1|t} = \{\psi_1\varepsilon_t + \psi_2\varepsilon_{t-1} + \cdots\}, \qquad \textit{Moving-Average Form}$$

$$\hat{y}_{t+1|t} = \{\pi_1 y_t + \pi_2 y_{t-1} + \cdots\}. \qquad \textit{Autoregressive Form}$$

Optimal Forecasts

Consider making a prediction at time t for h steps ahead. The value of the process at time $t+h$ is

$$\begin{aligned} y_{t+h} &= \{\psi_0\varepsilon_{t+h} + \psi_1\varepsilon_{t+h-1} + \cdots + \psi_{h-1}\varepsilon_{t+1}\} \\ &\quad + \{\psi_h\varepsilon_t + \psi_{h+1}\varepsilon_{t-1} + \cdots\}. \end{aligned} \tag{10}$$

The first term on the RHS embodies disturbances subsequent to the time t , and the second term embodies disturbances which are within the information set $\{\varepsilon_t, \varepsilon_{t-1}, \varepsilon_{t-2}, \ldots\}$. A linear forecasting function, based on the information set, takes the form of

$$\hat{y}_{t+h|t} = \{\rho_h\varepsilon_t + \rho_{h+1}\varepsilon_{t-1} + \cdots\}. \tag{11}$$

Then, given that $\varepsilon(t)$ is a white-noise process, it follows that the mean square of the error in the forecast h periods ahead is

$$E\big\{(y_{t+h} - \hat{y}_{t+h})^2\big\} = \sigma_\varepsilon^2 \sum_{i=0}^{h-1} \psi_i^2 + \sigma_\varepsilon^2 \sum_{i=h}^{\infty} (\psi_i - \rho_i)^2. \tag{12}$$

This is minimised by setting $\rho_i = \psi_i$; and so the optimal forecast is given by

$$\hat{y}_{t+h|t} = \{\psi_h\varepsilon_t + \psi_{h+1}\varepsilon_{t-1} + \cdots\}. \tag{13}$$

This might have been derived from the equation $y(t+h) = \psi(L)\varepsilon(t+h)$, which generates the true value of y_{t+h}, simply by putting zeros in place of the unobserved disturbances $\varepsilon_{t+1}, \varepsilon_{t+2}, \ldots, \varepsilon_{t+h}$ which lie in the future when the forecast is made.

On the assumption that the process is stationary, the mean-square error of the forecast tends to the value of

$$V\big\{y(t)\big\} = \sigma_\varepsilon^2 \sum \psi_i^2 \tag{14}$$

as the lead time h of the forecast increases. This is nothing but the variance of the process $y(t)$.

The optimal forecast of (5) may also be derived by specifying that the forecast error should be uncorrelated with the disturbances up to the time of making the forecast. For, if the forecast errors were correlated with some of the elements of the information set, then, as we have noted before, we would not be using the information efficiently, and we could not be generating optimal forecasts.

Generating The Forecasts Recursively

The optimal (m.m.s.e) forecast of y_{t+h} is the conditional expectation of y_{t+h} given the information set $\mathcal{I}_t$. Taking expectations conditional on $\mathcal{I}_t$ gives

$$\begin{aligned}&E(y_{t+k}|\mathcal{I}_t) = \hat{y}_{t+k|t} \quad \text{if} \quad k > 0,\\ &E(y_{t-j}|\mathcal{I}_t) = y_{t-j} \quad \text{if} \quad j \geq 0,\\ &E(\varepsilon_{t+k}|\mathcal{I}_t) = 0 \quad \text{if} \quad k > 0,\\ &E(\varepsilon_{t-j}|\mathcal{I}_t) = \varepsilon_{t-j} \quad \text{if} \quad j \geq 0.\end{aligned} \tag{21}$$

In this notation, the forecast h periods ahead is

$$\begin{aligned}E(y_{t+h}|\mathcal{I}_t) &= \sum_{k=1}^{h} \psi_{h-k} E(\varepsilon_{t+k}|\mathcal{I}_t) + \sum_{j=0}^{\infty} \psi_{h+j} E(\varepsilon_{t-j}|\mathcal{I}_t)\\ &= \sum_{j=0}^{\infty} \psi_{h+j} \varepsilon_{t-j}.\end{aligned} \tag{22}$$

The forecasts may be generated using a recursion based on the equation

$$\begin{aligned}y(t) = &-\big\{\alpha_1 y(t-1) + \alpha_2 y(t-2) + \cdots + \alpha_p y(t-p)\big\}\\ &+ \mu_0 \varepsilon(t) + \mu_1 \varepsilon(t-1) + \cdots + \mu_q \varepsilon(t-q).\end{aligned} \tag{23}$$

By taking the conditional expectation of this function, we get

$$\begin{aligned}\hat{y}_{t+h} = &-\{\alpha_1 \hat{y}_{t+h-1} + \cdots + \alpha_p y_{t+h-p}\}\\ &+ \mu_h \varepsilon_t + \cdots + \mu_q \varepsilon_{t+h-q} \quad \text{when} \quad 0 < h \leq p, q,\end{aligned} \tag{24}$$

$$\hat{y}_{t+h} = -\{\alpha_1 \hat{y}_{t+h-1} + \cdots + \alpha_p y_{t+h-p}\} \quad \text{if} \quad q < h \leq p, \tag{25}$$

$$\begin{aligned}\hat{y}_{t+h} = &-\{\alpha_1 \hat{y}_{t+h-1} + \cdots + \alpha_p \hat{y}_{t+h-p}\}\\ &+ \mu_h \varepsilon_t + \cdots + \mu_q \varepsilon_{t+h-q} \quad \text{if} \quad p < h \leq q,\end{aligned} \tag{26}$$

and

$$\hat{y}_{t+h} = -\{\alpha_1 \hat{y}_{t+h-1} + \cdots + \alpha_p \hat{y}_{t+h-p}\} \quad \text{when} \quad p, q < h. \tag{27}$$

It can be seen from (27) that, for $h > p, q$, the forecasting function becomes a pth-order homogeneous difference equation in y. The p values of $y(t)$ from $t = r = \max(p, q)$ to $t = r - p + 1$ serve as the starting values for the equation.

The Analytic Form of the Forecast Function

Beyond the reach of the starting values, the forecast function can be represented by a homogeneous difference equation. The unit roots can be incorporated within the analytic solution of the difference equation. In the long run, the unit roots dominate the solution.

In general, if d of the roots are unity, then the general solution will comprise a polynomial in t of order $d-1$.

Example. For an example of the analytic form of the forecast function, we may consider the Integrated Autoregressive (IAR) Process defined by

$$\big\{1-(1+\phi)L+\phi L^2\big\}y(t)=\varepsilon(t), \tag{30}$$

wherein $\phi \in (0,1)$. The roots of the auxiliary equation $z^2-(1+\phi)z+\phi=0$ are $z=1$ and $z=\phi$. The solution of the homogeneous difference equation

$$\big\{1-(1+\phi)L+\phi L^2\big\}\hat{y}(t+h|t)=0, \tag{31}$$

which defines the forecast function, is

$$\hat{y}(t+h|t)=c_1+c_2\phi^h, \tag{32}$$

where c_1 and c_2 are constants which reflect the initial conditions. These constants are found by solving the equations

$$\begin{aligned} y_{t-1}&=c_1+c_2\phi^{-1},\\ y_t&=c_1+c_2. \end{aligned} \tag{33}$$

The solutions are

$$c_1=\frac{y_t-\phi y_{t-1}}{1-\phi} \quad \text{and} \quad c_2=\frac{\phi}{\phi-1}(y_t-y_{t-1}). \tag{34}$$

The long-term forecast is $\bar{y}=c_1$ which is the asymptote to which the forecasts tend as the lead period h increases.

Ad-hoc Methods of Forecasting : The Holt–Winters Method

The Holt–Winters algorithm is useful in extrapolating local linear trends. The prediction h periods ahead of a series $y(t) = \{y_t, t = 0, \pm 1, \pm 2, \ldots\}$ which is made at time t is given by

$$\hat{y}_{t+h|t} = \hat{\alpha}_t + \hat{\beta}_t h, \tag{38}$$

where

$$\begin{aligned} \hat{\alpha}_t &= \lambda y_t + (1-\lambda)(\hat{\alpha}_{t-1} + \hat{\beta}_{t-1}) \\ &= \lambda y_t + (1-\lambda)\hat{y}_{t|t-1} \end{aligned} \tag{39}$$

is the estimate of an intercept or levels parameter formed at time t and

$$\hat{\beta}_t = \mu(\hat{\alpha}_t - \hat{\alpha}_{t-1}) + (1-\mu)\hat{\beta}_{t-1} \tag{40}$$

is the estimate of the slope parameter, likewise formed at time t. The coefficients $\lambda, \mu \in (0, 1]$ are the smoothing parameters.

The algorithm may also be expressed in error-correction form. Let

$$e_t = y_t - \hat{y}_{t|t-1} = y_t - \hat{\alpha}_{t-1} - \hat{\beta}_{t-1} \tag{41}$$

be the error at time t arising from the prediction of y_t on the basis of information available at time $t-1$. Then the formula for the levels parameter can be given as

$$\begin{aligned} \hat{\alpha}_t &= \lambda e_t + \hat{y}_{t|t-1} \\ &= \lambda e_t + \hat{\alpha}_{t-1} + \hat{\beta}_{t-1}, \end{aligned} \tag{42}$$

which, on rearranging, becomes

$$\hat{\alpha}_t - \hat{\alpha}_{t-1} = \lambda e_t + \hat{\beta}_{t-1}. \tag{43}$$

When the latter is drafted into equation (40), we get an analogous expression for the slope parameter:

$$\begin{aligned} \hat{\beta}_t &= \mu(\lambda e_t + \hat{\beta}_{t-1}) + (1-\mu)\hat{\beta}_{t-1} \\ &= \lambda\mu e_t + \hat{\beta}_{t-1}. \end{aligned} \tag{44}$$

The Holt–Winters Method and the IMA(2, 2) Model

In order reveal the underlying nature of the Holt–Winters method, it is helpful to combine the two equations (42) and (44) in a simple state-space model:

$$\begin{bmatrix} \hat{\alpha}(t) \\ \hat{\beta}(t) \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \hat{\alpha}(t-1) \\ \hat{\beta}(t-1) \end{bmatrix} + \begin{bmatrix} \lambda \\ \lambda\mu \end{bmatrix} e(t). \tag{45}$$

This can be rearranged to give

$$\begin{bmatrix} 1-L & -L \\ 0 & 1-L \end{bmatrix} \begin{bmatrix} \hat{\alpha}(t) \\ \hat{\beta}(t) \end{bmatrix} = \begin{bmatrix} \lambda \\ \lambda\mu \end{bmatrix} e(t). \tag{46}$$

The solution of the latter is

$$\begin{bmatrix} \hat{\alpha}(t) \\ \hat{\beta}(t) \end{bmatrix} = \frac{1}{(1-L)^2} \begin{bmatrix} 1-L & L \\ 0 & 1-L \end{bmatrix} \begin{bmatrix} \lambda \\ \lambda\mu \end{bmatrix} e(t). \tag{47}$$

Therefore, from (38), it follows that

$$\begin{aligned} \hat{y}(t+1|t) &= \hat{\alpha}(t) + \hat{\beta}(t) \\ &= \frac{(\lambda + \lambda\mu)e(t) + \lambda e(t-1)}{(1-L)^2}. \end{aligned} \tag{48}$$

This can be recognised as the forecasting function of an IMA(2, 2) model of the form

$$(I - L)^2 y(t) = \mu_0 \varepsilon(t) + \mu_1 \varepsilon(t-1) + \mu_2 \varepsilon(t-2) \tag{49}$$

for which

$$\hat{y}(t+1|t) = \frac{\mu_1 \varepsilon(t) + \mu_2 \varepsilon(t-1)}{(1-L)^2}. \tag{50}$$