5 Chapter 4: Forecasting with ARMA models

We consider forecasting using an ARMA(p, q) model defined as

$$\phi(B)X_t = \theta(B)\epsilon_t, \;\; \epsilon_t \sim WN(0, \sigma^2). \tag{5.1}$$

For forecasting, it is convenient to assume that the moving average operator $\theta(B)$ is invertible, i.e. the zeros of the corresponding polynomial $\theta(z)$ lie **strictly outside** (rather than merely "on or outside") the unit circle in the complex plane.

Recall that under the usual stationarity and invertibility conditions, it is possible to expand $C(z) = \theta(z)/\phi(z)$ as a power series $\sum_0^\infty c_r z^r$ to obtain the Wold representation $X_t = \sum_0^\infty c_r \epsilon_{t-r}$.

Under invertibility, we can similarly expand

$$D(z) = \frac{\phi(z)}{\theta(z)} = \sum_{r=0}^{\infty} d_r z^r \tag{5.2}$$

and so rewrite the ARMA model 5.1 as an infinite AR expansion

$$\epsilon_t = \sum_{r=0}^{\infty} d_r X_{t-r}. \tag{5.3}$$

The advantage of (5.3) is that, given the infinite past of the series $\{X_s, \;\; s < t\}$ we can solve to obtain exactly $\{\epsilon_s, \;\; s < t\}$.

Now suppose at time T, we are interested in predicting k steps ahead from available observations $\{X_t, \;\; t \leq T\}$ up to the time T. We denote such a k-steps ahead forecast $\hat{X}_{T,k}$. The **true unobserved** value at time $T + k$ is denoted by X_{T+k}.

Working with the Wold representation, we may consider forecasts of the form

$$\hat{X}_{T,k} = \sum_{r=0}^{\infty} c_{k,r}\epsilon_{T-r} \tag{5.4}$$

which, in view of the preceding discussion, is something that may be calculated, at least in theory, i.e. if we knew all ϵ_{T-r} for $r \in Z^+$.

Comparing (5.4) with the Wold representation for X_{T+k}, we see that

$$X_{T+k} - \hat{X}_{T,k} = \sum_{r=0}^{\infty} c_r\epsilon_{T+k-r} - \sum_{r=0}^{\infty} c_{k,r}\epsilon_{T-r} = \sum_{r=0}^{k-1} c_r\epsilon_{T+k-r} + \sum_{r=0}^{\infty}(c_{r+k} - c_{k,r})\epsilon_{T-r}. \tag{5.5}$$

Hence

$$E\left\{\left(X_{T+k} - \hat{X}_{T,k}\right)^2\right\} = \left\{\sum_{r=0}^{k-1} c_r^2 + \sum_{r=0}^{\infty}\left(c_{r+k} - c_{k,r}\right)^2\right\}\sigma_\epsilon^2. \tag{5.6}$$

This expression may be minimized by setting

$$c_{r,k} = c_{r+k}, \quad \text{for all} \;\; r \geq 0, \; k > 0, \tag{5.7}$$

and then gives rise to the mean squared prediction error

$$E\left\{\left(X_{T+k}-\hat{X}_{T,k}\right)^2\right\}=\left\{\sum_{r=0}^{k-1}c_r^2\right\}\sigma_\epsilon^2. \tag{5.8}$$

This therefore defines our theoretical optimal predictor, and its mean squared error. In practice, one does not usually go through the formality of constructing the Wold and infinite AR representations in this way. Instead, there is an alternative recursive approach, as follows.

Define $\hat{X}_{t,k}$ to be the optimal predictor of X_{t+k} given $X_1,\ldots,X_T$; for $-T+1\le k\le 0$, $\hat{X}_{T,k}=X_{T+k}$, i.e. for $k<0$ we do not need to forecast and can just take the observed value. We have the recursive relation

$$\hat{X}_{T,k}=\sum_{r=1}^{p}\phi_r\hat{X}_{T,k-r}+\hat{\epsilon}_{T+k}+\sum_{s=1}^{q}\theta_s\hat{\epsilon}_{T+k-s}. \tag{5.9}$$

For $k\le 0$, (5.9) allows us to calculate estimates of the one-step prediction errors, $\hat{\epsilon}_t$ for $1\le t\le T$. Then we apply (5.9) with $k>0$, defining $\hat{\epsilon}_t=0$ for $t>T$, to calculate the forecasts.

The difficulty that remains is how to start off the recursion (5.9). There are two standard solutions to that:

1. the *conditional* approach, in which we assume $X_t=\epsilon_t=0$ for all $t\le 0$,
2. the *backcasting* approach, in which we forecast the series in reverse direction to determine estimates of $X_0,\ X_{-1},\ldots$, as well as $\epsilon_0=\epsilon_{-1}=0$, etc.

An alternative superior approach is, however, to recast the model in state space form and apply the Kalman filter.

Another point to note throughout this discussion is that the "mean squared prediction errors" we have derived are based solely on the uncertainties of prediction: they do not make any allowance for errors in model identification. There exists an extended approach which also takes into account the standard errors of the parameter estimates.

Prediction of an AR(p) process. Applying (5.9) with $\theta_s=0$, for $s=1,2,\ldots,q$, we get

$$\hat{X}_{T;k}=\sum_{r=1}^{p}\psi_r\hat{X}_{T,k-r}=\psi_1\hat{X}_{T+k-1}+\psi_2\hat{X}_{T+k-2}+\ldots+\psi_p\hat{X}_{T+k-p}. \tag{5.10}$$

Prediction of an MA(q) process. Similarly applying (5.9) with $\psi_r=0$, for $r=1,2,\ldots,p$, we get

$$\hat{X}_{T,k}=\hat{\epsilon}_{T+k}+\sum_{s=1}^{q}\theta_s\hat{\epsilon}_{T+k-s}. \tag{5.11}$$

Example of forecasting from ARMA(p, q) in R

```
> X<-arima.sim(n = 1000, list(ar = c(0.5, -0.2),
ma = c(-0.4, 0.1)))
> fit <- arima(X, order = c(2,0,2))
> fit

Call: arima(x = X, order = c(2, 0, 2))

Coefficients:
         ar1      ar2      ma1     ma2  intercept
      0.6893  -0.3011  -0.6260  0.2017     0.0005
s.e.  0.2925   0.2311   0.3002  0.2370     0.0303

sigma^2 estimated as 1.035:  log likelihood = -1436,
aic = 2884.01

> predict(fit, n.ahead = 1)
$pred Time Series: Start = 1001 End = 1001
Frequency = 1 [1]
0.09895257

$se Time Series: Start = 1001 End = 1001
Frequency = 1 [1] 1.017201
```