Time Series Analysis – Lecture 9

Prediction

Dr. Kevin Burke

1 Introduction

Once we have found an appropriate model for an observed series, we typically wish to use this model to predict future values of the series (along with prediction limits).

Let $\mathcal{H}_t = \{Y_1, Y_2, \ldots, Y_{t-1}, Y_t\}$ be the **history** of the observed process up to time t, i.e., this encapsulates everything we know about the series at time t. The predicted value for Y_{t+1} *given* the history is

$$\hat{Y}_{t+1} = E(Y_{t+1} \,|\, \mathcal{H}_t)$$

and predicting l steps ahead gives

$$\hat{Y}_{t+l} = E(Y_{t+l} \,|\, \mathcal{H}_t)$$

where l is known as the **lead time**.

Note that

$$\hat{Y}_t = E(Y_t \,|\, \mathcal{H}_t) = Y_t$$

since, we *know* the value of Y_t at time t already so our prediction is just Y_t itself and, more generally, for $l \le 0$, $\hat{Y}_{t+l} = Y_{t+l}$.

2 AR

We will first consider prediction in the context of the autoregressive models. As prediction is more direct than in moving average models.

2.1 AR(1)

Consider an AR(1) model with non-zero mean

$$Y_t - \mu = \phi(Y_{t-1} - \mu) + e_t$$

which could also be written as $Y_t = \beta_0 + \phi Y_{t-1} + e_t$ but, for the purpose of prediction, it is easier to work with the model written as above.

We can then write

$$\begin{aligned} Y_{t+1} - \mu &= \phi(Y_t - \mu) + e_{t+1} \\ \Rightarrow Y_{t+1} &= \mu + \phi(Y_t - \mu) + e_{t+1}. \end{aligned}$$

Now, the taking the conditional expectation (conditional on $\mathcal{H}_t$) leads to the prediction:

$$\hat{Y}_{t+1} = \mu + \phi(\, E(Y_t \,|\, \mathcal{H}_t) - \mu) + E(e_{t+1} \,|\, \mathcal{H}_t)$$

Note the following:

- $E(Y_t \,|\, \mathcal{H}_t) = Y_t$ since Y_t is known at time t
- $E(e_{t+1} \,|\, \mathcal{H}_t) = E(e_{t+1}) = 0$ since e_{t+1} is independent of the past by definition and has expectation equal to zero.

Hence, the one-step ahead forecast is:

$$\hat{Y}_{t+1} = \mu + \phi(\, Y_t - \mu).$$

More generally, we can forecast l steps ahead as follows:

$$Y_{t+l} = \mu + \phi(Y_{t+l-1} - \mu) + e_{t+l}$$

$$\begin{aligned} \Rightarrow \hat{Y}_{t+l} &= E(Y_{t+l} \,|\, \mathcal{H}_t) \\ &= \mu + \phi(\, E(Y_{t+l-1} \,|\, \mathcal{H}_t) - \mu) + E(e_{t+l} \,|\, \mathcal{H}_t) \\ &= \mu + \phi(\hat{Y}_{t+l-1} - \mu). \end{aligned}$$

Thus, we can build up the forecasts as follows

$$\begin{aligned} \hat{Y}_{t+2} &= \mu + \phi(\hat{Y}_{t+1} - \mu) \\ &= \mu + \phi(\, \mu + \phi(\, Y_t - \mu) - \mu) \\ &= \mu + \phi^2(\, Y_t - \mu) \end{aligned}$$

$$\begin{aligned} \hat{Y}_{t+3} &= \mu + \phi(\hat{Y}_{t+2} - \mu) \\ &= \mu + \phi(\, \mu + \phi^2(\, Y_t - \mu) - \mu) \\ &= \mu + \phi^3(\, Y_t - \mu) \\ &\;\;\vdots \\ \hat{Y}_{t+l} &= \mu + \phi^l(\, Y_t - \mu). \end{aligned}$$

Note that, since $|\phi| < 0$,

$$\hat{Y}_{t+l} \approx \mu \qquad \text{for large } t,$$

i.e., **the long term forecast reverts to the process mean**. In fact this can be shown to be true for all ARMA models.

2.2 AR(2)

The AR(2) with non-zero mean is given by

$$Y_t - \mu = \phi_1(Y_{t-1} - \mu) + \phi_2(Y_{t-2} - \mu) + e_t.$$

Thus, replacing t with $t+1$ we get

$$Y_{t+1} = \mu + \phi_1(Y_t - \mu) + \phi_2(Y_{t-1} - \mu) + e_{t+1}$$

and, taking the conditional expectation leads to

$$\begin{aligned} \hat{Y}_{t+1} &= \mu + \phi_1(E(Y_t \,|\, \mathcal{H}_t) - \mu) + \phi_2(E(Y_{t-1} \,|\, \mathcal{H}_t) - \mu) \\ &\qquad + E(e_{t+1} \,|\, \mathcal{H}_t) \\ &= \mu + \phi_1(Y_t - \mu) + \phi_2(Y_{t-1} - \mu) \end{aligned}$$

since both Y_t and Y_{t-1} are known at time t. Similarly, it is easy to show that

$$\hat{Y}_{t+2} = \mu + \phi_1(\hat{Y}_{t+1} - \mu) + \phi_2(Y_t - \mu)$$

and, more generally,

$$\hat{Y}_{t+l} = \mu + \phi_1(\hat{Y}_{t+l-1} - \mu) + \phi_2(\hat{Y}_{t+l-2} - \mu).$$

3 MA

When making predictions from MA models, we need to learn how to evaluate

$$E(e_t \,|\, \mathcal{H}_t)$$

which does *not* simply become $E(e_t) = 0$ since e_t is not independent of $\mathcal{H}_t$.

Due to invertibility, we may write

$$\begin{aligned} e_t = Y_t - \pi_1 Y_{t-1} - \pi_2 Y_{t-2} - \cdots - \pi_{t-1} Y_1 \\ - \pi_t Y_0 - \pi_{t+1} Y_{-1} - \pi_{t+2} Y_{-2} - \cdots \end{aligned}$$

Furthermore, in order for the above sum converge, the π weights must be decaying. Hence, if t is large we can write

$$e_t \approx Y_t - \pi_1 Y_{t-1} - \pi_2 Y_{t-2} - \cdots - \pi_{t-1} Y_1$$

and, therefore,

$$\begin{aligned} &E(e_t \,|\, \mathcal{H}_t) \\ &\approx E(Y_t \,|\, \mathcal{H}_t) - \pi_1 E(Y_{t-1} \,|\, \mathcal{H}_t) - \cdots - \pi_{t-1} E(Y_1 \,|\, \mathcal{H}_t) \\ &= Y_t - \pi_1 Y_{t-1} - \cdots - \pi_{t-1} Y_1 \\ &\qquad\qquad\qquad \text{(since } Y_1, \ldots, Y_t \text{ are all known at time } t\text{)} \\ &\approx e_t. \end{aligned}$$

3.1 MA(1)

The MA model with non-zero mean is

$$Y_t = \mu + e_t - \theta e_{t-1}$$

Thus, replacing t with $t+1$ we get

$$Y_{t+1} = \mu + e_{t+1} - \theta e_{t-1}$$

and, taking the conditional expectation leads to

$$\begin{aligned} \hat{Y}_{t+1} &= \mu + E(e_{t+1} \,|\, \mathcal{H}_t) - \theta E(e_t \,|\, \mathcal{H}_t) \\ &\approx \mu - \theta e_t \end{aligned}$$

since, as before, e_{t+1} is independent of the past and $E(e_t \,|\, \mathcal{H}_t) \approx e_t$ as outlined above.

Note that we will have a value for e_t as a by-product of the estimating the parameters in the model. See Lecture 7 Section 4.2.

More generally, for $l > 1$,

$$\begin{aligned} \hat{Y}_{t+l} &= \mu + E(e_{t+l} \,|\, \mathcal{H}_t) - \theta E(e_{t+l-1} \,|\, \mathcal{H}_t) \\ &\approx \mu \end{aligned}$$

since both e_{t+l} and e_{t+l-1} are independent of the past when $l > 1$.

3.2 MA(2)

Consider the MA(2) process with non-zero mean:

$$\begin{aligned} Y_t &= \mu + e_t - \theta_1 e_{t-1} - \theta_1 e_{t-2} \\ \Rightarrow Y_{t+l} &= \mu + e_{t+l} - \theta_1 e_{t+l-1} - \theta_1 e_{t+l-2} \end{aligned}$$

Thus, applying the same steps as before leads to

$$\begin{aligned} \hat{Y}_{t+1} &= \mu - \theta_1 e_t - \theta_2 e_{t-1} \\ \hat{Y}_{t+2} &= \mu - \theta_2 e_t \\ \hat{Y}_{t+l} &= \mu \qquad\qquad \text{for } l > 2 \end{aligned}$$

Note: It is not strictly correct to write "=" (we should write "≈"). However, we will not worry about this.

4 ARMA

We can forecast from all ARMA models using the same ideas covered in the previous two sections.

4.1 ARMA(1,1)

Consider the ARMA(1,1) model with non-zero mean

$$\begin{aligned} Y_t - \mu &= \phi(Y_{t-1} - \mu) + e_t - \theta e_{t-1} \\ \Rightarrow Y_{t+l} &= \mu + \phi(Y_{t+l-1} - \mu) + e_{t+l} - \theta e_{t+l-1} \end{aligned}$$

It is straightforward to show that

$$\begin{aligned} \hat{Y}_{t+1} &= \mu + \phi(Y_t - \mu) - \theta e_t \\ \hat{Y}_{t+2} &= \mu + \phi(\hat{Y}_{t+1} - \mu) \\ &= \mu + \phi^2(Y_t - \mu) - \phi\,\theta e_t \\ &\vdots \\ \hat{Y}_{t+l} &= \mu + \phi^l(Y_t - \mu) - \phi^{l-1}\,\theta e_t \\ \Rightarrow \hat{Y}_{t+l} &\approx \mu \qquad \text{for large } l \end{aligned}$$

5 ARIMA

We now show that everything works in the same way for ARIMA models.

5.1 Random Walk

The random walk is defined as

$$Y_t = Y_{t-1} + e_t$$

$$\Rightarrow Y_{t+l} = Y_{t+l-1} + e_{t+l}$$

$$\begin{aligned}
\Rightarrow \hat{Y}_{t+1} &= Y_t \\
\hat{Y}_{t+2} &= \hat{Y}_{t+1} = Y_t \\
&\vdots \\
\hat{Y}_{t+l} &= \hat{Y}_{t+l-1} = Y_t \qquad \forall\, l
\end{aligned}$$

Thus, for the random walk, all forecasts are simply equal to the last observed value.

5.2 IMA(1,1)

The IMA(1,1) is defined as

$$\begin{aligned}
\nabla Y_t &= e_t - \theta e_{t-1} \\
Y_t &= Y_{t-1} + e_t - \theta e_{t-1}
\end{aligned}$$

$$\begin{aligned}
\Rightarrow \hat{Y}_{t+1} &= Y_t - \theta e_t \\
\hat{Y}_{t+2} &= \hat{Y}_{t+1} = Y_t - \theta e_t \\
&\vdots \\
\hat{Y}_{t+l} &= \hat{Y}_{t+l-1} = Y_t - \theta e_t \qquad \forall\, l
\end{aligned}$$

6 Forecast Error

The forecast error is given by

$$\hat{e}_{t+l} = Y_{t+l} - \hat{Y}_{t+l}$$

For all ARIMA models (details omitted) it turns out that

$$\hat{e}_{t+l} = \sum_{i=0}^{l-1} \psi_i \, e_{t+l-i}$$

From this we easily find that

$$E(\hat{e}_{t+l}) = \sum_{i=0}^{l-1} \psi_i \, E(e_{t+l-i}) = 0$$

which means **forecasts are unbiased**. Also, clearly

$$\begin{aligned}
\text{Var}(\hat{e}_{t+l}) &= \sum_{i=0}^{l-1} \psi_i^2 \, \text{Var}(e_{t+l-i}) \\
&\qquad \text{(since white noise terms are independent)} \\
&= \sum_{i=0}^{l-1} \psi_i^2 \, \sigma_e^2 \\
&= \sigma_e^2 \sum_{i=0}^{l-1} \psi_i^2
\end{aligned}$$

In particular note that $\text{Var}(\hat{e}_{t+1}) = \sigma_e^2$ since $\psi_0 = 1$. Recall from Lecture 4 that the ψ weights come from writing the model as a general linear process $Y_t = \psi(B)\, e_t = \sum_{i=0}^{\infty} \psi_i \, e_{t-i}$.

Although we will not prove the result $\hat{e}_{t+l} = \sum_{i=0}^{l-1} \psi_i \, e_{t+l-i}$, we will show that it holds for the AR(1) and MA(1) cases below.

6.1 AR(1)

From Section 2.1 the l step ahead forecast is given by

$$\hat{Y}_{t+l} = \mu + \phi^l (Y_t - \mu)$$

and, hence, the forecast error is

$$\begin{aligned}
\hat{e}_{t+l} &= Y_{t+l} - \hat{Y}_{t+l} \\
&= Y_{t+l} - \mu - \phi^l (Y_t - \mu).
\end{aligned}$$

In Lecture 4 (Section 4.4), we showed that the AR(1) process with zero-mean can be written as

$$Y_t = \sum_{i=0}^{\infty} \phi^i \, e_{t-i}$$

i.e., a general linear process with $\psi_i = \phi^i$. A non-zero mean can easily be incorporated by adding on μ:

$$Y_t = \mu + \sum_{i=0}^{\infty} \phi^i \, e_{t-i}$$

From this we get

$$Y_t - \mu = \sum_{i=0}^{\infty} \phi^i \, e_{t-i}$$

and

$$\begin{aligned}
Y_{t+l} - \mu &= \sum_{i=0}^{\infty} \phi^i \, e_{t+l-i} \\
&= \sum_{i=0}^{l-1} \phi^i \, e_{t+l-i} + \sum_{i=l}^{\infty} \phi^i \, e_{t+l-i} \\
&= \sum_{i=0}^{l-1} \phi^i \, e_{t+l-i} + \underbrace{\sum_{j=0}^{\infty} \phi^{j+l} \, e_{t-j}}_{\text{change of variables } j = i - l} \\
&= \sum_{i=0}^{l-1} \phi^i \, e_{t+l-i} + \phi^l \sum_{j=0}^{\infty} \phi^j \, e_{t-j} \\
&= \sum_{i=0}^{l-1} \phi^i \, e_{t+l-i} + \phi^l \, (Y_t - \mu)
\end{aligned}$$

Thus, the forecast error is

$$\begin{aligned}
\hat{e}_{t+l} &= Y_{t+l} - \mu - \phi^l (Y_t - \mu) \\
&= \sum_{i=0}^{l-1} \phi^i \, e_{t+l-i}
\end{aligned}$$

as required since $\psi_i = \phi^i$ for the AR(1) model.

Note that

$$\text{Var}(\hat{e}_{t+l}) = \sigma_e^2 \sum_{i=0}^{l-1} \phi^{2i}$$

6.2 MA(1)

We saw in Section 3.1 that

$$\hat{Y}_{t+1} = \mu - \theta e_t$$

$$\hat{Y}_{t+l} = \mu \qquad \text{for } l > 1$$

Since $Y_t = \mu + e_t - \theta e_{t-1}$, we easily find that

$$\hat{e}_{t+1} = e_t$$

and $\quad \hat{e}_{t+l} = -e_{t+l} + \theta e_{t+l-1} \qquad \text{for } l > 1$

which satisfies $\hat{e}_{t+l} = \sum_{i=0}^{l-1} \psi_i \, e_{t+l-i}$ as required since, for an MA(1) model, $\psi_0 = 1$, $\psi_1 = -\theta$ and $\psi_i = 0$ for $i > 1$.

Note that

$$\text{Var}(\hat{e}_{t+l}) = \begin{cases} \sigma_e^2 & \text{for } l = 1 \\ (1+\theta^2)\sigma_e^2 & \text{for } l > 1 \end{cases}$$

7 Prediction Limits

If the white noise terms are normally distributed then the forecast errors, which (as noted in Section 6) may be written as

$$\hat{e}_{t+l} = \sum_{i=0}^{l-1} \psi_i \, e_{t+l-i},$$

are also normally distributed (being a sum of normal variables).

Therefore, letting

$$s_{t+l} = \text{se}(\hat{e}_{t+l}) = \sqrt{\text{Var}(\hat{e}_{t+l})}$$

be the standard error of $\hat{e}_{t+l}$, we have

$$\Pr\left(-1.96 < \frac{\hat{e}_{t+l} - E(\hat{e}_{t+l})}{\text{se}(\hat{e}_{t+l})} < 1.96\right) = 0.95$$

and, therefore

$$\begin{aligned}
&\Pr\left(-1.96 < \frac{\hat{e}_{t+l} - 0}{s_{t+l}} < 1.96\right) \\
&= \Pr\left(-1.96 \, s_{t+l} < \hat{e}_{t+l} < 1.96 \, s_{t+l}\right) \\
&= \Pr\left(-1.96 \, s_{t+l} < Y_{t+l} - \hat{Y}_{t+l} < 1.96 \, s_{t+l}\right) \\
&= \Pr\left(\hat{Y}_{t+l} - 1.96 \, s_{t+l} < Y_{t+l} < \hat{Y}_{t+l} + 1.96 \, s_{t+l}\right) \\
&= 0.95
\end{aligned}$$

From this result we can be 95% confident that the true future value Y_{t+l} will be contained in the interval

$$\hat{Y}_{t+l} \pm 1.96 \, \text{se}(\hat{e}_{t+l})$$

$$\Rightarrow \hat{Y}_{t+l} \pm 1.96 \, \sigma_e \sqrt{\sum_{i=0}^{l-1} \psi_i^2}$$

For example

$$\hat{Y}_{t+1} \pm 1.96 \, \sigma_e$$

and

$$\hat{Y}_{t+2} \pm 1.96 \, \sigma_e \sqrt{1 + \psi_1^2}$$

are the 95% prediction limits for Y_{t+1} and Y_{t+2} respectively.