Forecasting

1. Optimal Forecast Criterion - Minimum Mean Square Error Forecast

We have now considered how to determine which ARIMA model we should fit to our data, we have also examined how to estimate the parameters in the ARIMA model and how to examine the goodness of fit of our model. So we are finally in a position to use our model to forecast new values for our time series. Suppose we have observed a time series up to time T, so that we have known values for $X_1, X_2, \cdots, X_T$, and subsequently, we also know $Z_1, Z_2, \cdots, Z_T$. This collection of information we know will be denoted as:

$$I_T = \{X_1, X_2, \cdots, X_T; Z_1, Z_2, \cdots, Z_T\}$$

Suppose that we want to forecast a value for X_{T+k}, k time units into the future. We now introduce the most prevalent optimality criterion for forecasting.

Theorem: The optimal k-step ahead forecast $X_{T+k,T}$ (which is a function of I_T) that will minimize the mean square error,

$$\mathrm{E}\left(X_{T+k} - X_{T+k,T}\right)^2$$

is

$$X_{T+k,T} = \mathrm{E}(X_{T+k} | I_T)$$

This optimal forecast is referred to as the **Minimum Mean Square Error Forecast.**

This optimal forecast is unbiased because

$$\mathrm{E}\left(X_{T+k} - X_{T+k,T}\right) = \mathrm{E}\left[\mathrm{E}\left(\left(X_{T+k} - X_{T+k,T}\right) | I_T\right)\right]$$

$$= \mathrm{E}\left[X_{T+k,T} - X_{T+k,T}\right] = 0$$

Lemma: Suppose Z and W are real random variables, then

$$\min_h E\left[\left(Z - h(W)\right)^2\right] = E\left[\left(Z - E(Z|W)\right)^2\right]$$

That is, the posterior mean E(Z|W) will minimize the quadratic loss (mean square error).

Proof:

$$E\left[\left(Z - h(W)\right)^2\right] = E\left[\left(Z - E(Z|W) + E(Z|W) - h(W)\right)^2\right]$$

$$= E\left[\left(Z - E(Z|W)\right)^2\right] + 2E\left[\left(Z - E(Z|W)\right)\left(E(Z|W) - h(W)\right)\right]$$

$$+E\left[\left(E(Z|W) - h(W)\right)^2\right]$$

Conditioning on W, the cross term is zero. Thus

$$E\left[\left(Z - h(W)\right)^2\right] = E\left[\left(Z - E(Z|W)\right)^2\right] + E\left[\left(E(Z|W) - h(W)\right)^2\right]$$

Note: We usually omit the word 'optimal' in the K Periods Ahead (Optimal) Forecast, because in the time series context, when we refer to 'forecast', we mean 'the optimal minimum mean square error forecast', by default.

2. Forecast in MA(1)

The MA(1) model is:

$$X_t = Z_t + \theta Z_{t-1}, \quad where\ Z_t \sim WN(0, \sigma^2)$$

Given the observed data $\{X_1, X_2, \cdots, X_T\}$, the white noise term Z_t is not directly observed, however, we can perform an ***Recursive Estimation of the White Noise*** as follows:

Given the initial condition Z_0 (not important), we have:

$$Z_1 = X_1 - \theta Z_0$$
$$Z_2 = X_2 - \theta Z_1$$
$$\cdots$$
$$Z_T = X_T - \theta Z_{T-1}$$

Therefore, given $\{X_1, X_2, \cdots, X_T\}$, we can also obtain $\{Z_1, Z_2, \cdots, Z_T\}$, the entire information known can thus be written as: $\boldsymbol{I_T = \{X_1, X_2, \cdots, X_T; Z_1, Z_2, \cdots, Z_T\}}$

Since $X_{T+1} = Z_{T+1} + \theta Z_T$, the **one period ahead (optimal) forecast** is

$$\begin{aligned} X_{T+1,T} &= E(X_{T+1}|I_T) = E(Z_{T+1} + \theta Z_T|I_T) \\ &= E(Z_{T+1}|I_T) + \theta E(Z_T|I_T) \\ &= E(Z_{T+1}) + \theta E(Z_T|Z_T) \\ &= 0 + \theta Z_T = \theta Z_T \end{aligned}$$

Note: Per convention in time series literature, we wrote:

$$E(Z_T|Z_T) = Z_T$$

Although one can write this more rigorously as:

$$E(Z_T|Z_T = z_T) = z_T$$

Similarly, we wrote

$$E(X_T|X_T) = X_T,$$

which can be written more rigorously as:

$$E(X_T|X_T = x_T) = x_T$$

The **forecast error** is

$$e_{T+1,T} = X_{T+1} - X_{T+1,T} = Z_{T+1}$$

Since the forecast is unbiased, the minimum mean square error is equal to the **forecast error variance**:

$$\begin{aligned} \mathrm{E}\left(X_{T+1} - X_{T+1,T}\right)^2 &= Var(X_{T+1} - X_{T+1,T}) = Var(Z_{T+1}) \\ &= \sigma^2 \end{aligned}$$

Given $I_T = \{X_1, X_2, \cdots, X_T; Z_1, Z_2, \cdots, Z_T\}$, and $X_{T+2} = Z_{T+2} + \theta Z_{T+1}$, the **<u>two periods ahead (optimal) forecast</u>** is

$$X_{T+2,T} = E(X_{T+2}|I_T) = E(Z_{T+2} + \theta Z_{T+1}|I_T)$$
$$= E(Z_{T+2}|I_T) + \theta E(Z_{T+1}|I_T)$$
$$= 0 + \theta * 0 = 0$$

The forecast error is

$$e_{T+2,T} = X_{T+2} - X_{T+2,T} = Z_{T+2} + \theta Z_{T+1}$$

Since the forecast is unbiased, the minimum mean square error is equal to the forecast error variance:

$$\mathrm{E}\left(X_{T+2} - X_{T+2,T}\right)^2 = Var(X_{T+2} - X_{T+2,T})$$
$$= Var(Z_{T+2} + \theta Z_{T+1}) = (1 + \theta^2)\sigma^2$$

In summary, for more than one period ahead, just like the two-periods ahead forecast, the forecast is zero, the unconditional mean. That is,

$$X_{T+2,T} = E(X_{T+2}|I_T) = 0$$
$$X_{T+3,T} = E(X_{T+3}|I_T) = 0$$
$$\cdots$$

The MA(1) process is not forecastable for more than one period ahead (apart from the unconditional mean).

Forecast for MA(1) with Intercept (non-zero mean)

If the MA(1) model includes an intercept

$$X_t = \mu + Z_t + \theta Z_{t-1}, \ \ where\ Z_t \sim WN(0, \sigma^2)$$

We can perform forecasting using the same approach.

For example, since $X_{T+1} = \mu + Z_{T+1} + \theta Z_T$, <u>the **one period ahead (optimal) forecast**</u> is

$$X_{T+1,T} = E(X_{T+1}|I_T) = E(\mu + Z_{T+1} + \theta Z_T|I_T)$$
$$= \mu + E(Z_{T+1}|I_T) + \theta E(Z_T|I_T)$$
$$= \mu + 0 + \theta Z_T = \mu + \theta Z_T$$

The forecast error is

$$e_{T+1,T} = X_{T+1} - X_{T+1,T} = Z_{T+1}$$

Since the forecast is unbiased, the minimum mean square error is equal to the forecast error variance:

$$\mathrm{E}\left(X_{T+1} - X_{T+1,T}\right)^2 = Var(X_{T+1} - X_{T+1,T}) = Var(Z_{T+1})$$
$$= \sigma^2$$

3. Generalization to MA(q):

For a MA(q) process, we can forecast up to q out-of-sample periods. Can you write down the (optimal) point forecast for each period, the forecast error, and the forecast error variance?

Please practice before the exam.

4. Forecast in AR(1)

For the AR(1) model,

$$X_t = \emptyset X_{t-1} + Z_t, where\ Z_t \sim WN(0, \sigma^2)$$

Given $I_T = \{X_1, X_2, \cdots, X_T; Z_1, Z_2, \cdots, Z_T\}$;

The **One Period Ahead (Optimal) Forecast** can be computed as follows:

$$X_{T+1} = \emptyset X_T + Z_{T+1}$$

So, given data I_T, the one period ahead forecast is

$$\begin{aligned} X_{T+1,T} &= E(X_{T+1}|I_T) = E(\emptyset X_T + Z_{T+1}|I_T) \\ &= E(\emptyset X_T|I_T) + E(Z_{T+1}|I_T) \\ &= \emptyset X_T + 0 \\ &= \emptyset X_T \end{aligned}$$

The **forecast error** is

$$e_{T+1,T} = X_{T+1} - X_{T+1,T} = Z_{T+1}$$

The **forecast error variance**:

$$Var(e_{T+1,T}) = Var(Z_{T+1}) = \sigma^2$$

Two Periods Ahead Forecast

By back-substitution

$$X_{T+2} = \emptyset X_{T+1} + Z_{T+2}$$
$$X_{T+2} = \emptyset(\emptyset X_T + Z_{T+1}) + Z_{T+2}$$
$$X_{T+2} = \emptyset^2 X_T + \emptyset Z_{T+1} + Z_{T+2}$$

So, given data I_T, the two periods ahead (optimal) forecast is

$$\begin{aligned} X_{T+2,T} &= E(X_{T+2}|I_T) = E(\emptyset^2 X_T + \emptyset Z_{T+1} + Z_{T+2}|I_T) \\ &= \emptyset^2 X_T + \emptyset 0 + 0 \\ &= \emptyset^2 X_T \end{aligned}$$

That is, the 2 periods ahead forecast is also a linear function of the final observed value, but with the coefficient $\emptyset^2$.

The forecast error is

$$e_{T+2,T} = X_{T+2} - X_{T+2,T} = \emptyset Z_{T+1} + Z_{T+2}$$

The forecast error variance is:

$$\mathrm{Var}(e_{T+2,T}) = Var(\emptyset Z_{T+1} + Z_{T+2}) = (1 + \emptyset^2)\sigma^2$$

K Periods Ahead Forecast

Similarly

$$X_{T+k} = \emptyset^k X_T + \emptyset^{k-1} Z_{T+1} + \cdots + \emptyset Z_{T+k-1} + Z_{T+k}$$

So, given I_T, the k periods ahead optimal forecast is

$$X_{T+k,T} = E(\mathrm{X}_{T+k}|\mathrm{I}_T) = \emptyset^k \mathrm{X}_T$$

AR(1) with Intercept

If the AR(1) model includes an intercept

$$\mathrm{X}_t = \alpha + \emptyset \mathrm{X}_{t-1} + Z_t, where\ Z_t \sim WN(0, \sigma^2)$$

Then the one period ahead forecast is

$$\begin{aligned} X_{T+1,T} &= E(X_{T+1}|I_T) \\ &= \alpha + E(\emptyset X_T|I_T) + E(Z_{T+1}|I_T) \\ &= \alpha + \emptyset X_T \end{aligned}$$

What is the two periods ahead forecast?

Forecast AR(1) recursively.

Alternatively, rather than using the back-substitution method directly as qwe have shown above, we can forecast the AR(1) recursively as follows:

$$\mathrm{X}_{T+2} = \emptyset \mathrm{X}_{T+1} + Z_{T+2}$$

So, given data I_T, the two periods ahead (optimal) forecast is

$$\begin{aligned} X_{T+2,T} &= E(X_{T+2}|I_T) = E(\emptyset \mathrm{X}_{T+1} + Z_{T+2}|I_T) = \emptyset E(\mathrm{X}_{T+1}|I_T) \\ &= \emptyset X_{T+1,T} \end{aligned}$$

Similarly, one can obtain the general recursive forecast relationship as:

$$\begin{aligned} X_{T+k,T} &= E(X_{T+k}|I_T) = E(\emptyset \mathrm{X}_{T+k-1} + Z_{T+k}|I_T) \\ &= \emptyset E(\mathrm{X}_{T+k-1}|I_T) = \emptyset X_{T+k-1,T} \\ &= \cdots = \emptyset^{k-1} X_{T+1,T} \end{aligned}$$

This way, once you have computed the one-step ahead forecast as:

$$\begin{aligned} X_{T+1,T} &= E(\mathrm{X}_{T+1}|I_T) = E(\emptyset \mathrm{X}_T + Z_{T+1}|I_T) \\ &= E(\emptyset \mathrm{X}_T|I_T) + E(Z_{T+1}|I_T) \\ &= \emptyset \mathrm{X}_T + 0 \\ &= \emptyset \mathrm{X}_T \end{aligned}$$

You can obtain the rest of the forecast easily.

Note: As mentioned in class, I prefer the back-substitution method (introduced first) in general because it can provide you with the forecast error directly as well.

5. AR(p) Process

Consider an AR(2) model with intercept

$$\mathrm{X}_t = \alpha + \emptyset_1 \mathrm{X}_{t-1} + \emptyset_2 \mathrm{X}_{t-2} + Z_t, where\ Z_t \sim WN(0, \sigma^2)$$

We have

$$X_{T+1} = \alpha + \emptyset_1 X_T + \emptyset_2 X_{T-1} + Z_{T+1}$$

Then the one period ahead forecast is

$$X_{T+1,T} = E(X_{T+1}|I_T)$$
$$= \alpha + E(\emptyset_1 X_T|I_T) + E(\emptyset_2 X_{T-1}|I_T) + E(Z_{T+1}|I_T)$$
$$= \alpha + \emptyset_1 X_T + \emptyset_2 X_{T-1}$$

The two periods ahead forecast (by the recursive foreecasting method) is

$$X_{T+2,T} = E(X_{T+2}|I_T)$$
$$= \alpha + E(\emptyset_1 X_{T+1}|I_T) + E(\emptyset_2 X_T|I_T) + E(Z_{T+2}|I_T)$$
$$= \alpha + \emptyset_1 X_{T+1,T} + \emptyset_2 X_T$$
$$= \alpha + \emptyset_1(\alpha + \emptyset_1 X_T + \emptyset_2 X_{T-1}) + \emptyset_2 X_T$$
$$= (1 + \emptyset_1)\alpha + \left(\emptyset_1{}^2 + \emptyset_2\right) X_T + \emptyset_1 \emptyset_2 X_{T-1}$$

Alternatively, the two periods ahead forecast (by the back-substitution method) can be derived as:

$$X_{T+2} = \alpha + \emptyset_1 X_{T+1} + \emptyset_2 X_T + Z_{T+2}$$
$$= \alpha + \emptyset_1(\alpha + \emptyset_1 X_T + \emptyset_2 X_{T-1} + Z_{T+1}) + \emptyset_2 X_T + Z_{T+2}$$
$$= (1 + \emptyset_1)\alpha + \left(\emptyset_1{}^2 + \emptyset_2\right) X_T + \emptyset_1 \emptyset_2 X_{T-1}$$
$$+ \emptyset_1 Z_{T+1} + Z_{T+2}$$

Therefore

$$X_{T+2,T} = E(X_{T+2}|I_T)$$
$$= (1 + \emptyset_1)\alpha + \left(\emptyset_1{}^2 + \emptyset_2\right) X_T + \emptyset_1 \emptyset_2 X_{T-1}$$

The forecast error is

$$e_{T+2,T} = X_{T+2} - X_{T+2,T} = \emptyset_1 Z_{T+1} + Z_{T+2}$$

The forecast error variance is:

$$\mathrm{Var}(e_{T+2,T}) = Var(\emptyset_1 Z_{T+1} + Z_{T+2}) = (1 + \emptyset_1^2)\sigma^2$$

What are the forecasts for more future periods?
Please practice before the exam.

6. ARMA Process

Consider an ARMA(1,1) model

$$X_t = \emptyset X_{t-1} + Z_t + \theta Z_{t-1}, where\ Z_t \sim WN(0, \sigma^2)$$

Again we can use the back-substitution method to compute point forecast.

$$X_{T+1} = \emptyset X_T + Z_{T+1} + \theta Z_T$$
$$X_{T+2} = \emptyset X_{T+1} + Z_{T+2} + \theta Z_{T+1}$$
$$= \emptyset(\emptyset X_T + Z_{T+1} + \theta Z_T) + Z_{T+2} + \theta Z_{T+1}$$

One-step ahead forecast:

$$X_{T+1,T} = E(X_{T+1}|I_T)$$
$$= E(\emptyset X_T|I_T) + E(Z_{T+1}|I_T) + E(\theta Z_T|I_T)$$
$$= \emptyset X_T + \theta Z_T$$

Two-step ahead forecast:

$$X_{T+2,T} = E(X_{T+2}|I_T)$$
$$= E(\emptyset X_{T+1}|I_T) + E(Z_{T+2}|I_T) + E(\theta Z_{T+1}|I_T)$$
$$= \emptyset E(X_{T+1}|I_T) = \emptyset(\emptyset X_T + \theta Z_T)$$

The two-step ahead forecast error is

$$e_{T+2,T} = X_{T+2} - X_{T+2,T} = \emptyset Z_{T+1} + Z_{T+2} + \theta Z_{T+1}$$
$$= (\emptyset + \theta)Z_{T+1} + Z_{T+2}$$

Again, since the forecast is unbiased, the minimum mean square error is equal to the forecast error variance:

$$\mathrm{E}\left(X_{T+2} - X_{T+2,T}\right)^2 = Var(X_{T+2} - X_{T+2,T})$$
$$= Var[(\emptyset + \theta)Z_{T+1} + Z_{T+2}]$$
$$= [(\emptyset + \theta)^2 + 1]\sigma^2$$

7. ARIMA Process

Suppose X_t follows the ARIMA(1,1,0) model. That means its first difference

$$Y_t = \nabla X_t = X_t - X_{t-1}$$

follows the ARMA(1,0) model, which is simply the AR(1) model as follows:

$$Y_t = \phi Y_{t-1} + Z_t$$
$$Here\ we\ assume\ Z_t \sim WN(0, \sigma_Z^2)$$

Note: we used σ_Z^2 instead of σ^2 so you can get familiar with different notations.

Substituting $Y_t = \nabla X_t = X_t - X_{t-1}$, and $Y_{t-1} = \nabla X_{t-1} = X_{t-1} - X_{t-2}$ we can write the original ARIMA(1,1,0) model as:

$$X_t - X_{t-1} = \phi(X_{t-1} - X_{t-2}) + Z_t$$

That is:

$$X_t = (1 + \phi)X_{t-1} - \phi X_{t-2} + Z_t$$

One-step ahead forecast:

$$X_{T+1,T} = E[X_{T+1}|I_T] = E[(1+\phi)X_T - \phi X_{T-1} + Z_{T+1}|I_T]$$
$$= (1+\phi)X_T - \phi X_{T-1}$$

Thus the forecast error

$$e_{T+1,T} = X_{T+1} - X_{T+1,T}$$
$$= (1+\phi)X_T - \phi X_{T-1} + Z_{T+1} - [(1+\phi)X_T - \phi X_{T-1}]$$
$$= Z_{T+1}$$

And the variance of the forecast error

$$Var\left(e_{T+1,T}\right) = Var(Z_{T+1}) = \sigma_Z^2$$

Two-step ahead forecast:

$$X_{T+2} = (1+\phi)X_{T+1} - \phi X_T + Z_{T+2}$$
$$= (1+\phi)[(1+\phi)X_T - \phi X_{T-1} + Z_{T+1}] - \phi X_T + Z_{T+2}$$
$$= (1+\phi+\phi^2)X_T - (\phi+\phi^2)X_{T-1} + Z_{T+2} + (1+\phi)Z_{T+1}$$

Therefore

$$X_{T+2,T} = E[X_{T+2}|I_T]$$
$$= E[(1+\phi+\phi^2)X_T - (\phi+\phi^2)X_{T-1} + Z_{T+2} + (1+\phi)Z_{T+1}|I_T]$$
$$= (1+\phi+\phi^2)X_T - (\phi+\phi^2)X_{T-1}$$

Thus the forecast error

$$e_{T+2,T} = X_{T+2} - X_{T+2,T} = Z_{T+2} + (1+\phi)Z_{T+1}$$

And the variance of the forecast error

$$Var\left(e_{T+2,T}\right) = (2+2\phi+\phi^2)\sigma_Z^2$$

8. Prediction (Forecast) Error, and Prediction Interval

Recall the *k*-step ahead **prediction error** is defined as

$$\mathrm{e}_{\mathrm{T+k,T}} = \mathrm{X}_{\mathrm{T+k}} - \mathrm{E}(\mathrm{X}_{\mathrm{T+k}}|\mathrm{I}_{\mathrm{T}})$$

By the law of total variance:

$$\mathrm{Var(Y) = E[Var(Y|X)] + Var[E(Y|X)]}$$

We can derive the **variance of the prediction error:**

$$\mathrm{Var}\left(\mathrm{e}_{\mathrm{T+k,T}}\right) = \mathrm{Var}(\mathrm{X}_{\mathrm{T+k}}|\mathrm{I}_{\mathrm{T}})$$

Based on the predictor and its error variance, we may construct the prediction intervals. For example, if the X_t's are normally distributed, we can consider the 95% prediction interval

$$\mathrm{E}(\mathrm{X}_{\mathrm{T+k}}|\mathrm{I}_{\mathrm{T}}) \pm \mathrm{Z}_{0.025}\sqrt{\mathrm{Var}(\mathrm{X}_{\mathrm{T+k}}|\mathrm{I}_{\mathrm{T}})}$$


ECONOMIC
FORECAST
THE CHRISTIAN SCIENCE MONITOR