Forecasting cycles: AR( ), MA( ), and ARMA( )

Dr. Natalia Khorunzhina

CBS

Information Set

- To forecast y_{T+h} at time T we use relevant information.
- Most information is values of the economic variable.
- Those observed up to time T.
- This includes current and past values of the variable y_t.
- It can also include other relevant variables.
- All this information is the *Information Set*, written as Ω_T.
- For example $\Omega_T = \{y_1, y_2, y_3, \ldots, y_T\}$ is the set of current and previous values.
- $\Omega_T = \{y_1, x_1, y_2, x_2, y_3, x_3, \ldots, y_T, x_T\}$ includes another variable x_t.

Information Set

- So long as y is covariance stationary, for both MA(q) and AR(p) processes we can express the information set in terms of current and past shocks:

$$\Omega_T = (\varepsilon_1, \varepsilon_2, ..., \varepsilon_{T-1}, \varepsilon_T)$$

- Suppose that the process to be forecasted is a covariance stationary AR(1): $y_t = \beta y_{t-1} + \varepsilon_t$.
- Then, immediately

$$\begin{aligned} \varepsilon_T &= y_T - \beta y_{T-1} \\ \varepsilon_{T-1} &= y_{T-1} - \beta y_{T-2} \\ \varepsilon_{T-2} &= y_{T-2} - \beta y_{T-3} \end{aligned}$$

- We can view the information set as containing the current and past values of y and ε:

$$\Omega_T = (y_1, y_2, ..., y_{T-1}, y_T, \varepsilon_1, \varepsilon_2, ..., \varepsilon_{T-1}, \varepsilon_T)$$

Information Set

- Based on the information set, we want to find **optimal forecast** of y at some future time.
- The optimal forecast is the **conditional mean** $E(y_{T+h}|\Omega_T)$, the expected value of the future value of the series being forecast, conditional on available information.
- Our forecasting method will be the same: we write out the process for the future time period of interest, $T+h$ and condition it on what is known up to time T.

Constructing forecast for MA(1)

- If the error is unforecastable $E(\varepsilon_t \mid \Omega_{t-1}) = 0$ for any period t then the conditional mean of y_{T+1} is

$$\begin{aligned} E(y_{T+1} \mid \Omega_T) &= E(\varepsilon_{T+1} + \theta\varepsilon_T \mid \Omega_T) \\ &= E(\varepsilon_{T+1} \mid \Omega_T) + \theta E(\varepsilon_T \mid \Omega_T) \\ &= \theta\varepsilon_T \end{aligned}$$

- This is the best forecast of y_{T+1}.
- The optimal forecast error is

$$\begin{aligned} y_{T+1} - E(y_{T+1} \mid \Omega_T) &= (\varepsilon_{T+1} + \theta\varepsilon_T) - \theta\varepsilon_T \\ &= \varepsilon_{T+1} \end{aligned}$$

Conditional Variance of MA(1)

- The conditional variance of y_{T+1} is

$$\begin{aligned} var(y_{T+1} \mid \Omega_T) &= var(y_{T+1} - E(y_{T+1} \mid \Omega_T) \mid \Omega_T) \\ &= var(\varepsilon_{T+1} \mid \Omega_T) \\ &= \sigma^2 \end{aligned}$$

- The conditional variance, the forecast variance, and the innovation variance are all the same thing

Example: MA(2) Process

$$
\begin{aligned}
y_t &= \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} \\
\varepsilon_t &\sim WN(0, \sigma^2)
\end{aligned}
$$

- Suppose we are standing at time T and we want to forecast y_{T+h}.
- Let's write out the process for $T+1$:

$$
y_{T+1} = \varepsilon_{T+1} + \theta_1 \varepsilon_T + \theta_2 \varepsilon_{T-1}
$$

- All we know at time $T+1$ is the values from the information set: $\Omega_T = (y_1, y_2, ..., y_{T-1}, y_T, \varepsilon_1, \varepsilon_2, ..., \varepsilon_{T-1}, \varepsilon_T)$

Example: MA(2) Process

- Condition y_{T+1} on time T information set:

$$\begin{aligned} E(y_{T+1}|\Omega_T) &= E(\varepsilon_{T+1}|\Omega_T) + \theta_1 E(\varepsilon_T|\Omega_T) + \theta_2 E(\varepsilon_{T-1}|\Omega_T) \\ &= 0 + \theta_1 \varepsilon_T + \theta_2 \varepsilon_{T-1} \\ &= \theta_1 \varepsilon_T + \theta_2 \varepsilon_{T-1} \end{aligned}$$

- To forecast 2 steps ahead, apply time T information set for the process y_{T+2}:

$$\begin{aligned} E(y_{T+2}|\Omega_T) &= E(\varepsilon_{T+2}|\Omega_T) + \theta_1 E(\varepsilon_{T+1}|\Omega_T) + \theta_2 E(\varepsilon_T|\Omega_T) \\ &= 0 + \theta_1 \times 0 + \theta_2 \varepsilon_T \\ &= \theta_2 \varepsilon_T \end{aligned}$$

- To forecast 3 steps ahead, continue in this fashion:

$$\begin{aligned} E(y_{T+3}|\Omega_T) &= E(\varepsilon_{T+3}|\Omega_T) + \theta_1 E(\varepsilon_{T+2}|\Omega_T) + \theta_2 E(\varepsilon_{T+1}|\Omega_T) \\ &= 0 + \theta_1 \times 0 + \theta_2 \times 0 \\ &= 0 \end{aligned}$$

Example: MA(2) Process

Let's compute the corresponding **forecast errors**:

$$\begin{aligned}
e_{T+1} &= \varepsilon_{T+1} \\
e_{T+2} &= \varepsilon_{T+2} + \theta_1 \varepsilon_{T+1} \\
e_{T+3} &= \varepsilon_{T+3} + \theta_1 \varepsilon_{T+2} + \theta_2 \varepsilon_{T+1}
\end{aligned}$$

The forecast error variances are:

$$\begin{aligned}
\sigma_1^2 &= \sigma^2 \\
\sigma_2^2 &= \sigma^2(1 + \theta_1^2) \\
\sigma_3^2 &= \sigma^2(1 + \theta_1^2 + \theta_2^2)
\end{aligned}$$

It appears that MA(q) process is not forecastable (apart from the unconditional mean) more that q steps ahead. Also, the forecast intervals are rapidly increasing, suggesting less precision to the forecast.

Constructing Interval Forecast

- We assume that the regression disturbances are normally distributed.
- A 95% interval forecast ignoring parameters estimation uncertainty is $\hat{y}_{T+h} \pm 1.96\hat{\sigma}_h$.
- Consider again an MA(2) process.

$$\begin{aligned} y_t &= \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} \\ \varepsilon_t &\sim WN(0, \sigma^2) \end{aligned}$$

  - Assuming normality, the one-step ahead interval forecast is $\hat{y}_{T+1} \pm 1.96\hat{\sigma} = (\hat{\theta}_1 e_T + \hat{\theta}_2 e_{T-1}) \pm 1.96\hat{\sigma}$.
  - Assuming normality, the two-step ahead interval forecast is $\hat{y}_{T+2} \pm 1.96\hat{\sigma}\sqrt{1+\hat{\theta}_1^2} = \hat{\theta}_2 e_T \pm 1.96\hat{\sigma}\sqrt{1+\hat{\theta}_1^2}$.

Constructing forecast for AR(1)

- Conditional mean:

$$E(y_{T+1} \mid \Omega_T) = E(\beta y_T + \varepsilon_{T+1} \mid \Omega_T) = \beta y_T$$

- Conditional variance:

$$\begin{aligned} var(y_{T+1} \mid \Omega_T) &= var(y_{T+1} - E(y_{T+1} \mid \Omega_T) \mid \Omega_T) \\ &= var(\varepsilon_{T+1} \mid \Omega_T) \\ &= \sigma^2 \end{aligned}$$

Forecast of AR(1) process

- A very simple recursive method for computing the optimal forecast is available in the autoregressive case.
- Consider $AR(1)$ process $y_t = \beta y_{t-1} + \varepsilon_t$
- To construct the 1-step-ahead forecast, we write out the process for time $T+1$,

$$y_{T+1} = \beta y_T + \varepsilon_{T+1}$$

- Then, projecting on time T information set, we obtain $E(y_{T+1} \mid \Omega_T) = \beta y_T$
- Now, let's consider the 2-step-ahead forecast. Write down the process for time $T+2$: $y_{T+2} = \beta y_{T+1} + \varepsilon_{T+2}$
- Then, projecting on time T information set, we obtain

$$E(y_{T+2} \mid \Omega_T) = \beta E(y_{T+1} \mid \Omega_T)$$
$$E(y_{T+2} \mid \Omega_T) = \beta^2 y_T$$

Forecast of ARMA(1,1) process

- We write out the $ARMA(1,1)$ process at time $T+1$:

$$y_{T+1} = \beta y_T + \varepsilon_{T+1} + \theta \varepsilon_T$$

- Projecting on the time T information set yields:

$$E(y_{T+1} \mid \Omega_T) = \beta y_T + \theta \varepsilon_T$$

- Now, let's find $E(y_{T+2} \mid \Omega_T)$. The process at time $T+2$ is:

$$y_{T+2} = \beta y_{T+1} + \varepsilon_{T+2} + \theta \varepsilon_{T+1}$$

- Projecting on the info set:

$$E(y_{T+2} \mid \Omega_T) = \beta E(y_{T+1} \mid \Omega_T)$$

- Substituting, we obtain: $E(y_{T+2} \mid \Omega_T) = \beta^2 y_T + \beta\theta\varepsilon_T$
- Continuing, we have $E(y_{T+h} \mid \Omega_T) = \beta E(y_{T+h-1} \mid \Omega_T)$, for all $h > 1$

Training and test sets

- The accuracy of forecasts can also be determined by how well a model performs on new data that were not used by the model.
- We can separate the data in 2 parts: **training** and **test** data;
- The training data is used to estimate a model and the test data is used to evaluate the accuracy of the forecast;
- The size of the test set is typically about 10-20% of the total sample.

Measures of the forecast accuracy

- Compute the forecast $\hat{y}_{T+h}$ and the forecast errors on the test set:

$$e_{T+h} = y_{T+h} - \hat{y}_{T+h}$$

- Most commonly used measures are mean error (ME), mean absolute error (MAE) and root mean squared error (RMSE):

$$ME = mean(e_t), \qquad t = T+1, ..., T+h$$

$$MAE = mean(|e_t|), \qquad t = T+1, ..., T+h$$

$$RMSE = \sqrt{mean(e_t^2)}, \qquad t = T+1, ..., T+h$$

- The model with the smallest MAE or RMSE and with the ME closest to 0 provides the closest prediction of the actual data.

Percentage-based measure of the forecast accuracy

- To compare forecast performances between data sets (for example, the forecasts on log-transformed and untransformed data), compute percentage errors

$$p_t = 100e_t/y_t$$

- They have an advantage of being unit-free
- Most commonly used measure is mean absolute percentage error (MAPE):

$$MAPE = mean(|p_t|), \qquad t = T+1, ..., T+h$$

- The model with the smallest MAPE provides the closest prediction of the actual data.