Lecture Notes 6:

Forecasting in MA, AR, and ARMA

ECMT 475: Economic Forecasting (Spring 2011)

Guangyi Ma
Optimal Point Forecasts

- Textbook Chapter 9

- Conditional mean as the optimal point forecast:

\[ y_{T+s,T} = E(y_{T+s} | \Omega_T). \]

- Minimizing expected loss.

- The Forecasts made by using MA, AR, or ARMA models are linear forecast.
Optimal Forecast in MA(1)

- In the MA(1) model
  \[ y_t = \varepsilon_t + \theta \varepsilon_{t-1} \text{ where } \varepsilon_t \sim WN \left(0, \sigma^2\right), \]

- Given data \( \{y_1, y_2, \cdots, y_T\} \), the one period ahead **optimal forecast** is
  \[
  y_{T+1,T} = E (y_{T+1}|\Omega_T) \\
  = E (\varepsilon_{T+1}|\Omega_T) + \theta E (\varepsilon_T|\Omega_T) \\
  = 0 + \theta \varepsilon_T = \theta \varepsilon_T. 
  \]

- The **forecast error** is
  \[
  y_{T+1} - y_{T+1,T} = \varepsilon_{T+1}. 
  \]
• The forecast error variance is

\[ \text{Var} \left( y_{T+1} - y_{T+1, T} \right) = \text{Var} \left( \varepsilon_{T+1} \right) = \sigma^2. \]
Recursive Forecast for MA(1)

- The white noise term $\varepsilon_T$ is not directly observed.

- In practice, we can numerically compute the errors by a recursive approach.

- Given the initial condition $\varepsilon_0 = 0$ (not important),

\[
\begin{align*}
\hat{\varepsilon}_1 &= y_1 - \hat{\theta}\varepsilon_0, \\
\hat{\varepsilon}_2 &= y_2 - \hat{\theta}\hat{\varepsilon}_1, \\
&\vdots \\
\hat{\varepsilon}_T &= y_T - \hat{\theta}\hat{\varepsilon}_{T-1},
\end{align*}
\]
• The one period ahead forecast is \( \hat{y}_{T+1,T} = \hat{\theta} \hat{\varepsilon}_T \).

• For more than one period ahead, the forecast is zero, the unconditional mean. (Why?)

\[
\hat{y}_{T+2,T} = E \left( y_{T+2} | \Omega_T \right) = 0 \\
\hat{y}_{T+3,T} = E \left( y_{T+3} | \Omega_T \right) = 0 \\
\ldots
\]

• That is, the MA(1) process is not forecastable more than one period ahead (apart from the unconditional mean).
Recursive Forecast for MA(1) with Intercept

- If the MA(1) model includes an intercept

\[ y_t = \mu + \varepsilon_t + \theta \varepsilon_{t-1} \text{ where } \varepsilon_t \sim WN(0, \sigma^2) \]

- We can still use the recursive approach.

- Given the initial condition \( \varepsilon_0 = 0 \),

\[
\begin{align*}
\hat{\varepsilon}_1 & = y_1 - \hat{y}_1 = y_1 - (\hat{\mu} + \hat{\theta} \varepsilon_0), \\
\hat{\varepsilon}_2 & = y_2 - \hat{y}_2 = y_2 - (\hat{\mu} + \hat{\theta} \varepsilon_1), \\
\vdots & \\
\hat{\varepsilon}_T & = y_T - \hat{y}_T = y_T - (\hat{\mu} + \hat{\theta} \varepsilon_{T-1}),
\end{align*}
\]
The one period ahead forecast is \( \hat{y}_{T+1,T} = \hat{\mu} + \hat{\theta} \hat{\varepsilon}_T \).

For more than one period ahead, the forecast is the (non-zero) unconditional mean:

\[
\begin{align*}
\hat{y}_{T+2,T} &= \hat{\mu} \\
\hat{y}_{T+3,T} &= \hat{\mu} \\
\cdots
\end{align*}
\]

**Generalization**: For a MA(q) process, we can forecast up to q out-of-sample periods. Can you write the point forecast for each period, the forecast error, and the forecast error variance?
Optimal Forecast in AR(1)

- In the AR(1) model
  \[ y_t = \beta y_{t-1} + \epsilon_t \text{ where } \epsilon_t \sim WN(0, \sigma^2) \]

- Given data \( \{y_1, y_2, \cdots, y_T\} \), the one period ahead optimal forecast is
  \[
y_{T+1,T} = E(y_{T+1}|\Omega_T) \\
  = E(\beta y_T|\Omega_T) + E(\epsilon_{T+1}|\Omega_T) \\
  = \beta y_T + 0 = \beta y_T
  \]

- The one period ahead forecast is a linear function of the final observed value.
Two Periods Ahead Forecast

- By back-substitution

\[ y_t = \beta y_{t-1} + \epsilon_t \]
\[ = \beta (\beta y_{t-2} + \epsilon_{t-1}) + \epsilon_t \]
\[ = \beta^2 y_{t-2} + \beta \epsilon_{t-1} + \epsilon_t \]

so

\[ E(y_t|\Omega_{t-2}) = E\left(\beta^2 y_{t-2} + \beta \epsilon_{t-1} + \epsilon_t|\Omega_{t-2}\right) \]
\[ = \beta^2 y_{t-2} + \beta \times 0 + 0 \]
\[ = \beta^2 y_{t-2} \]
• Given data \( \{y_1, y_2, \cdots, y_T\} \), the two periods ahead \textbf{optimal forecast} is

\[
y_{T+2, T} = E(y_{T+2}|\Omega_T) = \beta^2 y_T
\]

• That is, the 2 periods ahead forecast is also a linear function of the final observed value, but with the coefficient \( \beta^2 \).
K Periods Ahead Forecast

- Similarly

\[ y_t = \beta^k y_{t-k} + \beta^{k-1} \epsilon_{t-k+1} + \cdots + \beta \epsilon_{t-1} + \epsilon_t \]

so

\[ E(y_t|\Omega_{t-k}) = \beta^k y_{t-k} \]

- Given data \( \{y_1, y_2, \cdots, y_T\} \), the \( k \) periods ahead optimal forecast is

\[ y_{T+k,T} = E(y_{T+k}|\Omega_T) = \beta^k y_T \]
AR(1) with Intercept

- If the AR(1) model includes an intercept
  \[ y_t = \alpha + \beta y_{t-1} + \epsilon_t \quad \text{where} \quad \epsilon_t \sim WN \left(0, \sigma^2 \right) \]

- Then the one period ahead forecast is
  \[ y_{T+1,T} = E(y_{T+1}|\Omega_T) = \alpha + E(\beta y_T|\Omega_T) + E(\epsilon_{T+1}|\Omega_T) = \alpha + \beta y_T \]

- In practice, we compute \( \hat{y}_{T+1,T} = \hat{\alpha} + \hat{\beta} y_T \) using the estimates.

- What is the two period ahead forecast?
Forecast Error

• Same as in the MA(1) case, the one period ahead forecast error of AR(1) is

\[ y_{T+1} - \hat{y}_{T+1,T} = \varepsilon_{T+1} \]

• The forecast error variance is

\[ Var\left(y_{T+1} - \hat{y}_{T+1,T}\right) = Var\left(\varepsilon_{T+1}\right) = \sigma^2 \]

• In STATA, use "predict yhat, xb" to compute point forecast \( \hat{y}_{T+1,T} \).

• Use "predict ymse, mse" to compute forecast error variance \( \hat{\sigma}^2 \).
Interval Forecast

- Use the Normal method:
  - Assume forecast error is normally distributed
  - Confidence interval constructed by
    \[
    \left[ \hat{y}_{T+1,T} - \hat{\sigma} \cdot z_{1-\alpha/2}, \hat{y}_{T+1,T} + \hat{\sigma} \cdot z_{1-\alpha/2} \right]
    \]
  - Note that the critical value \( z_{1-\alpha/2} = -z_{\alpha/2} \).

- For a 95% interval (\( \alpha = 0.05 \)): \[
\left[ \hat{y}_{T+1,T} - 1.96\hat{\sigma}, \hat{y}_{T+1,T} + 1.96\hat{\sigma} \right]
\]

- For a 90% interval (\( \alpha = 0.10 \)): \[
\left[ \hat{y}_{T+1,T} - 1.645\hat{\sigma}, \hat{y}_{T+1,T} + 1.645\hat{\sigma} \right]
\]
• "gen c1 = yhat - 1.96*sqrt(ymse)" to compute lower bound

• label var c1 "95% lower bound"

• "gen c2 = yhat + 1.96*sqrt(ymse)" to compute lower bound

• label var c2 "95% upper bound"

• "twoway (tsline y) (tsline yhat) (tsline c1) (tsline c2)" to plot the forecast results

• Tips: When you do not have the data for the periods to be forecast (as in a real forecast project!), use "tsappend, add(12)" to generate 12 periods as desired before computing point forecast values.
AR(p) Process

- Consider an AR(2) model with intercept

\[ y_t = \alpha + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \epsilon_t \text{ where } \epsilon_t \sim WN \left( 0, \sigma^2 \right) \]

- Then the one period ahead forecast is

\[
y_{T+1, T} = E \left( y_{T+1} | \Omega_T \right) \\
= \alpha + E \left( \beta_1 y_T | \Omega_T \right) + E \left( \beta_2 y_{T-1} | \Omega_T \right) + E \left( \epsilon_{T+1} | \Omega_T \right) \\
= \alpha + \beta_1 y_T + \beta_2 y_{T-1}
\]
The two period ahead forecast is

\[ y_{T+2,T} = E(y_{T+2}|\Omega_T) \]
\[ = \alpha + E(\beta_1 y_{T+1}|\Omega_T) + E(\beta_2 y_T|\Omega_T) + E(\varepsilon_{T+2}|\Omega_T) \]
\[ = \alpha + \beta_1 y_{T+1,T} + \beta_2 y_T \]
\[ = \alpha + \beta_1 (\alpha + \beta_1 y_T + \beta_2 y_{T-1}) + \beta_2 y_T \]
\[ = (1 + \beta_1) \alpha + (\beta_1^2 + \beta_2) y_T + \beta_1 \beta_2 y_{T-1} \]

What are the forecasts for more future periods?

Here we used the recursive method again.

This is called the chain rule of forecasting in the textbook.
Consider an ARMA(1,1) model

\[ y_t = \alpha + \beta y_{t-1} + \epsilon_t + \theta \epsilon_{t-1} \] where \( \epsilon_t \sim WN \left(0, \sigma^2\right) \)

• Again we can use the recursive method to compute point forecast.

• Might be cumbersome to write, but STATA can do it for you!