Forecasting based on lagged $\varepsilon' s$:

Consider an MA(∞) process:

$y_t - \mu = \psi(L)\varepsilon_t, \qquad \varepsilon_t \sim WN(0, \sigma^2)$

where

$\psi(L) = \sum_{j=0}^{\infty} \psi_j L^j$

$\psi_0 = 1$

$\sum_{j=0}^{\infty} |\psi_j| < \infty$

- Suppose we observe an infinite history of ε_t up to date $t : \{\varepsilon_t, \varepsilon_{t-1}, \varepsilon_{t-2}, ...\}$
- Also suppose we know the MA parameters $\mu\{\psi_j\}_{j=0}^{\infty}$

Then,

$y_{t+s} = \mu + \varepsilon_{t+s} + \psi_1\varepsilon_{t+s-1} + \ldots + \psi_{s-1}\varepsilon_{t+1} + \psi_s\varepsilon_t + \psi_{s+1}\varepsilon_{t+1} + \ldots$

The optimal forecast of y_{t+s} in terms of MSE is:

$E[y_{t+s}|\varepsilon_t, \varepsilon_{t-1}, \ldots] = \mu + \psi_s\varepsilon_t + \psi_{s+1}\varepsilon_{t-1} + \ldots$

Note: This is different from $y_t = \mu + \psi_0\varepsilon_t + \psi_1\varepsilon_{t-1} + \ldots$

The forecast error is:

$y_{t-s} - E[y_{t+s}|\varepsilon_t, \varepsilon_{t-1}, \ldots] = \mu + \overbrace{\varepsilon_{t+s} + \psi_1\varepsilon_{t+s-1} + \psi_2\varepsilon_{t+s-2} + \ldots} + \psi_s\varepsilon_t + \psi_{s+1}\varepsilon_{t+1} + \ldots - \mu - \psi_s\varepsilon_t - \psi_{s+1}\varepsilon_{t-1} - \ldots \implies = \varepsilon_{t+s} + \psi_1\varepsilon_{t+s-1} + \ldots + \psi_{s-1}\varepsilon_{t+1}$

Since $E[y_{t+s}|\varepsilon_t, \varepsilon_{t-1}, \ldots]$ is linear in $\{\varepsilon_\tau\}_{\tau=-\infty}^{\infty}$ it is both the optimal forecast and optimal linear forecast.

- Hamilton refers to optimal linear forecasts as $\hat{E}[y_{t+s}|\varepsilon_t, \varepsilon_{t-1}, \ldots]$
- But in this case $E[y_{t+s}|\varepsilon_t, \ldots] = \hat{E}[y_{t+s}|\varepsilon_t, \ldots] \implies y^*_{t+s|t} = \hat{y}_{t+s|t}$ which is also a linear projection

  $\hat{p}(y_{t+s}|\varepsilon_t, \varepsilon_{t-1}, \ldots]$

Clearly, the linear projection condition is satisfied for $j = t, t-1, \ldots$

$E[(y_{t+s} - E[y_{t+s}|\varepsilon_t, \varepsilon_{t-1}, \ldots])\varepsilon_j] = E[(\varepsilon_{t+s} + \psi_1\varepsilon_{t+s-1} + \ldots + \psi_{s-1}\varepsilon_{t+1})\varepsilon_j] = 0$

The forecast MSE is:

$E[(y_{t+s} - E[y_{t+s}|\varepsilon_t, \varepsilon_{t-1}, \ldots])^2] = E[(\varepsilon_{t+s} + \psi_1\varepsilon_{t+s-1} + \ldots + \psi_{s-1}\varepsilon_{t+1})^2] = \sigma^2 \sum_{j=0}^{s-1} \psi_j^2$

Forecasting based on lagged y's:

Suppose we don't observe the full history of ε_t, bu instead observe the full history of $y_t : y_t, y_{t-1}, y_{t-2}, \ldots$

Given the same $MA(\infty)$ process as before:

$y_t - \mu = \psi(L)\varepsilon_t, \qquad \varepsilon_t \sim WN(0, \sigma^2)$

If the $MA(\infty)$ representation is invertible, we can write it as an $AR(\infty)$:

$\eta(L)(y_t - \mu) = \varepsilon_t$

where $\eta(L) = \psi^{-1}(L)$

The history of ε_t can be constructed with the history of y_t.

$\varepsilon_t = \eta(L)(y_t - \mu)$

$\varepsilon_{t-1} = \eta(L)(y_{t-1} - \mu)$

$\varepsilon_{t-2} = \eta(L)(y_{t-2} - \mu)$

$\vdots$

So,

$E[y_{t+s}|\varepsilon_t, \varepsilon_{t-1}, \ldots] = E[y_{t+s}|y_t, y_{t-1}, \ldots] = \mu + (\psi_s + \psi_{s+1}L + \psi_{s+2}L^2 + \ldots)\varepsilon_t$

$= \mu + (\psi_s + \psi_{s+1}L + \psi_{s+2}L^2 + \ldots)\eta(L)(y_t - \mu) = \mu + (\psi_s + \psi_{s+1}L + \psi_{s+2}L^2 + \ldots) \cdot ((y_t - \mu) - \eta_1(y_{t-1} - \mu) - \eta_2(y_{t-2} - \mu) - \ldots)$

Ex. AR(1)

For an AR(1) with $|\phi| < 1$:

$y_t - \mu = \psi(L)\varepsilon_t$

where

$\psi(L) = (1 + \phi(L) + \phi^2 L^2 + \ldots) = (1 + \psi_1 + \psi_2 + \ldots)$

The optimal forecast s-periods ahead is

$E[y_{t+s}|\varepsilon_t, \varepsilon_{t-1}, \ldots] = \mu + \psi_s\varepsilon_t + \psi_{s+1}\varepsilon_{t-1} + \ldots = \mu + \phi^s\varepsilon_t + \phi^{s+1}\varepsilon_{t-1} + \ldots$

$= \mu + \phi^s(\varepsilon_t + \phi\varepsilon_{t-1} + \phi^2\varepsilon_{t-2} + ...) = \mu + \phi^s(y_t - \mu)$

The forecast decays toward μ as s increases.

The MSE is

$MSE = \sigma^2 \sum_{j=0}^{s-1} psi_j^2 = \sigma^2 \sum_{j=0}^{s-1} \phi^{2j}$

As $s \to \infty, MSE \to \frac{\sigma^2}{1-\phi^2} = var(y_t)$

Forecasts based on a finite number of observations

In reality, we don't observe an infinite history of $y_t, y_{t-1}, y_{t-2}, \ldots$

Suppose we have only a finite set of m past observations of $y_t : y_t, y_{t-1}, \ldots, y_{t-m+1}$

- The optimal AR(p) forecast only makes use of the past p observations if available (i.e. $p < m$)
- But if we want to forecast an MA or ARMA (of any orders), we need an infinite history to construct an optimal forecast.

Approximate optimal forecasts

Start by setting all ε's prior to time t-m+1 equal to zero.

$E[y_{t+s}|y_t, y_{t-1}, \ldots] \approx E[y_{t+s}|y_t, y_{t-1}, \ldots, y_{t-m+1}, \varepsilon_{t-m} = 0, \varepsilon_{t-m-1} = 0, \ldots]$

MA(q)

Start with

$\hat{\varepsilon}_{t-m} = \hat{\varepsilon}_{t-m-1} = \ldots = \hat{\varepsilon}_{t-m-q+1} = 0$

Calculate forward recursively

$\hat{\varepsilon}_{t-m+1} = (y_{t-m+1} - \mu)$

$\hat{\varepsilon}_{t-m+2} = (y_{t-m+2} - \mu) - \theta_1 \hat{\varepsilon}_{t-m+1}$

$\hat{\varepsilon}_{t-m+3} = (y_{t-m+3} - \mu) - \theta_1 \hat{\varepsilon}_{t-m+2} - \theta_2 \hat{\varepsilon}_{t-m+1}$

$\vdots$

With $\hat{\varepsilon}_t, \hat{\varepsilon}_{t-1}, \ldots, \hat{\varepsilon}_{t-m+1}$ in hand we can compute forecasts

$\hat{y}_{t+s} = \theta_s \hat{\varepsilon}_t + \theta_{s+1} \hat{\varepsilon}_{t-1} + \ldots + \theta_q \hat{\varepsilon}_{t-q+s}$

<u>Exact Finite Sample Forecasts</u>

Another forecast approximation method is to simply project $y_{t+1} - \mu$ on $\vec{X}_t = \begin{bmatrix} y_t - \mu \\ . \\ . \\ . \\ y_{t-m+1} - \mu \end{bmatrix}$

That is $\hat{y}^{(m)}_{t+1|t} - \mu = \vec{X}_t' \vec{\beta}^{(m)} = \beta_1^{(m)}(y_t - \mu) + \beta_2^{(m)}(y_{t-1} - \mu) + \ldots + \beta_m^{(m)}(y_{t-m+1} - \mu)$

$$\vec{\beta}^{(m)} = E[\vec{X}_t \vec{X}_t']^{-1} E[\vec{X}_t y_{t+1}^{-m}] = \begin{bmatrix} \gamma_0 & \gamma_1 & \gamma_2 & \cdots & \gamma_{m-1} \\ \gamma_1 & \gamma_0 & \gamma_1 & \cdots & \gamma_{m-2} \\ \gamma_2 & \gamma_1 & \gamma_0 & \cdots & \gamma_{m-3} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \gamma_{m-1} & \cdots & \cdots & \cdots & \gamma_0 \end{bmatrix}^{-1} \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_m \end{bmatrix}$$

Similarly, $y^{(m)}_{t+s|t} - \mu = \beta_1^{(m,s)}(y_t - \mu) + \beta_2^{(m,s)}(y_{t-1} - \mu) + \ldots + \beta_m^{(m,s)}(y_{t-m+1} - \mu) = \vec{X}_t' \vec{\beta}^{(m,s)}$

As before, $\vec{\beta}^{(m,s)} = E[\vec{X}_t \vec{X}_t']^{-1} E[\vec{X}_t y_{t+s}^{-m}] = \begin{bmatrix} \gamma_0 & \gamma_1 & \gamma_2 & \cdots & \gamma_{m-1} \\ \gamma_1 & \gamma_0 & \gamma_1 & \cdots & \gamma_{m-2} \\ \gamma_2 & \gamma_1 & \gamma_0 & \cdots & \gamma_{m-3} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \gamma_{m-1} & \cdots & \cdots & \cdots & \gamma_0 \end{bmatrix}^{-1} \begin{bmatrix} \gamma_s \\ \gamma_{s+1} \\ \vdots \\ \gamma_{s+m-1} \end{bmatrix}$

Ex: ARMA(1,1)

Suppose $|\phi| < 1$ and $|\theta| < 1 \rightarrow$ causal and invertible. Then:

$(1 - \phi L)(y_t - \mu) = (1 + \theta L)\varepsilon_t$

So,

$y_t - \mu = \psi(L)\varepsilon_t$

where $\quad \psi(L) = (1 - \psi L)^{-1}(1 + \theta L)\varepsilon_t$

We can also write

$\varepsilon_t = (1+\theta L)^{-1}(1-\phi L)(y_t - \mu) = \psi(L)^{-1}(y_t - \mu)$.

We can write

$\psi(L) = (1 + \phi L + \phi^2 L^2 + \ldots)(1+\theta L) = 1 + (\phi+\theta)L + \phi^2 + \phi\theta)L^2 + (\phi^3 + \phi^2\theta)L^3 + \ldots$

$\psi(L) = 1 + \sum_{j=1}^{\infty}(\phi^j + \phi^{j-1}\theta)L^j$

$\implies \psi_m = \phi^m + \phi^{m-1}\theta$

Let's define $\psi_s(L)$ as the polynomial

$\psi_s(L) = \psi_s + \psi_{s+1}L + \psi_{s+2}L^2 + \ldots$

This is different from

$\psi_s L^s + \psi_{s+1}L^{s+1} + \ldots$

For the ARMA(1,1),

$\psi_s(L) = (\phi^s + \phi^{s-1}\theta) + (\phi^{s+1} + \phi^s\theta)L + (\phi^{s+2} + \phi^{s+1}\theta)L^2 + \ldots = \sum_{j=2}^{\infty}(\phi^j + \phi^{j-1}\theta)L^{j-s}$

$= (\phi^s + \phi^{s-1}\theta)\sum_{j=0}^{\infty}\phi^j L^j = (\phi^s + \phi^{s-1}\theta)(1-\phi L)^{-1}$

Recall, for an $MA(\infty)$, the optimal forecast is

$\hat{y}_{t+s|t} - \mu = E[y_{t+s}|\varepsilon_t, \varepsilon_{t-1}, \ldots] = \psi_s\varepsilon_t + \psi_{s+1}\varepsilon_{t-1} + \psi_{s+2}\varepsilon_{t-2} + \ldots = \psi_s(L)\varepsilon_t$

So, for the ARMA(1,1)

$\hat{y}_{t+s|t} - \mu = (\phi^s + \phi^{s-1}\theta)(1-\phi L)^{-1}\varepsilon_t = (\phi^s + \phi^{s-1}\theta)(1-\phi L)^{-1}(1-\phi L)(1+\theta L)^{-1}(y_t - \mu)$

$= (\phi^s + \phi^{s-1}\theta)(1+\theta L)^{-1}(y_t - \mu)$

Note:

$\hat{y}_{t+s|t} - \mu = (\phi^s + \phi^{s-1}\theta)(1+\theta L)^{-1}(y_t - \mu)$

$\hat{y}_{t+s|t} - \mu = \phi(\phi^{s-1} + \phi^{s-2}\theta)(1+\theta L)^{-1}(y_t - \mu)$

$\hat{y}_{t+s|t} - \mu = \phi(\hat{y}_{t+s-1|t} - \mu), \quad \text{if } s \geq 2$

which means, the forecast decays toward μ,

Foor s = 1,

$\hat{y}_{t+s|t} - \mu = (\phi+\theta)(1+\theta L)^{-1}(y_t - \mu) = (\phi + \phi\theta L - \phi\theta L + \theta)(1+\theta L)^{-1}(y_t - \mu)$

$= [\phi(1+\theta L) + \theta(1-\phi L)](1+\theta L)^{-1}(y_t - \mu)$

$= \phi(y_t - \mu) + \theta(1-\phi L)(1+\theta L)^{-1}(y_t - \mu) = \phi(y_t - \mu) + \theta\varepsilon_t$