Conditional forecasting in ARMA models

(Updated Spring 2021)

Chaoyi Chen
Institute of MNB, Corvinus University of Budapest

Empirical Financial Econometrics

Outlines

- Conditional forecasting in ARMA models ⏩ Jump [Self-study if you have interest :)]
- Constructing out of sample forecasts ⏩ Jump [Online Lecture]

Conditional forecasting in ARMA models

- AR(1) forecast
- AR(2) forecast
- MA(1) forecast
- MA(2) forecast
- ARMA(1,1) forecast
- ARMA(p,q) forecast

One and multi-period ahead forecasts

- Let k = forecast horizon
- At time t:
  - know y_t
  - forecast y_{t+k} - k period ahead forecast
    - * $k = 1$: 1 period forecast
      e.g. daily data: forecast Monday's value on Sunday
    - * $k > 1$: multi-period forecast
      e.g. daily data: forecast Friday's value on Sunday
- General principle
  - Date conditional expectation by time you form forecast:
    $E_t \longrightarrow$ forecast made at time $t \longrightarrow$ No variables dated $t+1, t+2, ...$ can be used in forecast (don't know there yet)
  - The variable you want to forecast is dated $t+k$
    - * want to forecast: y_{t+k}
  - Putting this together, my forecast of y_{t+h} at time t is $E_t y_{t+k}$

1). AR(1)

$$y_t = a_0 + a_1 y_{t-1} + \varepsilon_t;\ E_{t-1}\varepsilon_t = 0$$

- 1-period forecast:

$$y_{t+1} = a_0 + a_1 y_t + \varepsilon_t$$

$$E_t y_{t+1} = a_1 + a_1 \times \underbrace{E_t y_t}_{\text{known at time } t} + \underbrace{E_t[\varepsilon_{t+1}]}_{=0 \text{ by assumption}}$$

$$E_t y_{t+1} = a_0 + a_1 y_t \text{ infeasible forecast}$$

$$\widehat{y}_{t+1} = a_0 + a_1 y_t \text{ feasible forecast}$$

Specific models: AR(1) - 2 period forecast

- 2-period forecast: $(E_t y_{t+2})$

$$y_{t+2} = a_0 + a_1 y_{t+1} + \varepsilon_{t+2} \tag{1}$$
$$E_t y_{t+2} = a_0 + a_1 E_t(\underbrace{y_{t+1}}_{\text{RV at time } t}) + E_t \varepsilon_{t+2}$$

- We need the useful property - Law of Iterative Expectations (L.I.E)
  LIE detail
- We use L.I.E to show that $E_t \varepsilon_{t+2} = 0$

$$E_t(\varepsilon_{t+2}) \underset{\text{by LIE}}{=} E_t\{\underbrace{E_{t+1}(\varepsilon_{t+2})}_{=0 \text{ by assumption}}\} = E_t[0] = 0$$

Specific models: AR(1) - 2 period forecast cont.

- Plugging this result back into (6) gives us:

$$E_t y_{t+2} = a_0 + a_1 E_t[y_{t+1}] + \underbrace{E_t \varepsilon_{t+2}}_{=0} \quad (2)$$

$$E_t y_{t+2} = a_0 + a_1 E_t[y_{t+1}] \text{ (infeasible)}$$

- Now substitute $E_t[y_{t+1}] = a_0 + a_1 y_t$ into (2), we have

$$E_t y_{t+2} = a_0 + a_1[a_0 + a_1 y_2] \quad (3)$$

$$E_t y_{t+2} = a_0(1 + a_1) + a_1^2 y_t \text{ (infeasible)}$$

- (2) and (3) are infeasible forecasts
- The feasible analog to (2) is

$$\widehat{y}_{t+2} = \widehat{a}_0 + \widehat{a}_1 \widehat{y}_{t+1}; \ \widehat{y}_{t+1} = \widehat{a}_0 + \widehat{a}_1 y_t \quad (4)$$

  which is convenient if we want to forecast both y_{t+1} and y_{t+2}. The feasible analog to (3) is

$$\widehat{y}_{t+2} = \widehat{a}_0(1 + \widehat{a}_1) + \widehat{a}_1^2 y_t \quad (5)$$

Specific models: AR(1) - 3 period forecast

- 3-period forecast:

$$y_{t+3} = a_0 + a_1 y_{t+2} + \varepsilon_{t+2}$$
$$E_t y_{t+3} = a_0 + a_1 E_t y_{t+2} + E_t \varepsilon_{t+3}$$
$$\text{note that } E_t \varepsilon_{t+3} \underset{\text{by LIE}}{=} E_t \{ \underbrace{E_{t+2} \varepsilon_{t+3}}_{=0 \text{ by assumption}} \} = 0$$

$$E_t y_{t+3} = a_0 + a_1 E_t y_{t+2} \text{ (infeasible)} \tag{6a}$$
$$\widehat{y}_{t+3} = a_0 + a_1 \widehat{y}_{t+2} \text{ (feasible)} \tag{6b}$$

Specific models: AR(1) - 3 period forecast cont.

- OR plug in (3) for $E_t y_{t+2} = a_0(1 + a_1) + a_1^2 y_t$ to get

$$E_t y_{t+3} = a_0 + a_1 E_t y_{t+2}$$
$$= a_0 + a_1\{a_0(1 + a_1) + a_1^2 y_t\}$$
$$= a_0 + a_0(a_1 + a_1^2) + a_1^3 y_t$$

$$E_t y_{t+3} = a_0(1 + a_1 + a_1^2) + a_1^3 y_t \quad (7a)$$

which is made feasible by $\widehat{y}_{t+3} = \widehat{a}_0(1 + \widehat{a}_1 + \widehat{a}_1^2) + \widehat{a}_1^3 y_t$ (7b)

Specific models: AR(1) - K period forecast

- Now we can observe the pattern: Extending (6a) and (6b)

$$E_t y_{t+k} = a_0 + a_1 E_t y_{t+k-1} \tag{8a}$$

$$\widehat{y}_{t+k} = \widehat{a}_0 + \widehat{a}_1 \widehat{y}_{t+k-1} \tag{8b}$$

(8b) is easy to implement if you want to forecast all of $y_{t+1}, y_{t+2}, ..., y_{t+k}$

- <u>OR</u> extending (7a) and (7b)

$$E_t y_{t+k} = a_0(1 + a_1 + a_1^2 + ... + a_1^{k-1}) + a_1^k y_t = a_0 \sum_{i=0}^{k-1} a_1^i + a_1^k y_t \tag{9a}$$

$$\widehat{y}_{t+k} = \widehat{a}_0 \sum_{i=0}^{k-1} \widehat{a}_1^i + \widehat{a}_1^k y_t \tag{9b}$$

Compare to unconditional forecast

- Recall: Unconditional forecast made using unconditional expectation ($E[y_t]$)
- So, solve for $E[y_t]$ assuming stationarity

$$\begin{aligned} y_t &= a_0 + a_1 y_{t-1} + \varepsilon_t \\ E[y_t] &= a_0 + a_1 E[\underbrace{y_{t-1}}_{\text{still random at time } -\infty}] + E[\varepsilon_t] \end{aligned}$$

Using the L.I.E,

$$E[\varepsilon_t] \underset{\text{by L.I.E}}{=} E\{\underbrace{E_{t-1}(\varepsilon_t)}_{=0 \text{ by assumption}}\} = 0 \Longrightarrow E[y_t] = a_0 + a_1 E[y_{t-1}]$$

$$\text{By stationarity: } E[y_t] = E[y_{t-1}] \Longrightarrow E[y_t] = a_0 + a_1 E[y_t]$$

$$\Longrightarrow E[y_t] = a_1 E[y_t] = a_0 \Longrightarrow (1 - a_1 E[y_t]) = a_0$$

$$\boxed{E[y_t] = \frac{a_0}{1 - a_1}} \tag{10a}$$

$$\text{With feasible version: } \frac{\widehat{a}_0}{1 - \widehat{a}_1} \text{ or simply } \bar{y} \tag{10b}$$

Compare to unconditional forecast cont.

- Now compare (10) to (9)
- From (9), $E_t y_{t+k} = a_0 \sum_{i=0}^{k-1} a_1^i + a_1^k y_t$.
- What happens for large k?

$$\text{Becasue } —a_1 < 1— \text{ stationarity} \implies a_1^k \longrightarrow 0 \text{ as } k \longrightarrow \infty$$

$$\implies \lim_{k \longrightarrow \infty} E_t y_{t+k} = a_0 \sum_{i=0}^{\infty} a_1^i$$

$$= a_0 \frac{1}{1 - a_1} = \frac{a_0}{1 - a_1} = E[y_{t+k}]$$

$$\implies \boxed{\lim_{k \longrightarrow \infty} E_t y_{t+k} = E[y_{t+k}]} \tag{11}$$

- As forecast horizon increases our conditional forecast is more approaching to our unconditional forecast
- This is because the information we condition on only useful for predicting near future

Compare to unconditional forecast intuition

- In other words, in there is a reversion to unconditional mean in long run
- If today's y_t is about mean, we expect tomorrow & next day's y_{t+1} & y_{t+2} also to be above mean. Conditional forecast gives this
- But in 40 years from now, we expect y_{t+k} to revert back to mean
- In other words, y_t has almost no information about y_{t+40} years. So conditioning on y_t doesn't help & we might as well use unconditional forecast Plot

Specific models: AR(2) - 1 period forecast

2). AR(2)

$$y_t = a_0 + a_1 y_{t-1} + a_2 y_{t-2} + \varepsilon_t; \;\; E_{t-1}\varepsilon_t = 0$$

- <u>1 period ahead forecast</u>

$$y_{t+1} = a_0 + a_1 y_t + a_2 y_{t-1} + \varepsilon_{t+1}$$
$$\Longrightarrow E_t y_{t+1} = a_0 + a_1 E_t[\underbrace{y_t}_{\text{known at t}}] + a_2 E_t[\underbrace{y_{t-1}}_{\text{known at t}}] + \underbrace{E_t[\varepsilon_{t+1}]}_{=0}$$

$$\Longrightarrow E_t y_{t+1} = a_0 + a_1 y_t + a_2 y_{t-1} \tag{12a}$$
$$\text{Feasible version}: \;\; \widehat{y}_{t+1} = \widehat{a}_0 + \widehat{a}_1 y_t + \widehat{a}_2 y_{t-1} \tag{12b}$$

Specific models: AR(2) - 2 period forecast

- <u>2 period ahead forecast</u>

$$y_{t+2} = a_0 + a_1 y_{t+1} + a_2 y_t + \varepsilon_{t+2}$$
$$\implies E_t y_{t+2} = a_0 a_1 E_t[\underbrace{y_{t+1}}_{\text{still random at t}}] + a_2 E_t[\underbrace{y_t}_{\text{known at t}}] + E_t[\underbrace{\varepsilon_{t+2}}_{=0 \text{ by LIE}}]$$

$$\longrightarrow E_t y_{t+2} = a_0 + a_1 E_t[y_{t+1}] + a_2 y_t \tag{13a}$$
$$\text{Feasible version}: \ \widehat{y}_{t+2} = \widehat{a}_0 = \widehat{a}_1 \widehat{y}_{t+1} + \widehat{a}_2 y_t \tag{13b}$$

- We can substitute in for $E_t y_{t+1}$ but this gets tedious, so I'll skip this and let you try it at home :)

Specific models: AR(2) - 3 period forecast

- <u>3 period ahead forecast</u>

$$y_{t+3} = a_0 + a_1 y_{t+2} + a_2 y_{t+1} + \varepsilon_{t+3}$$

$$E_t y_{t+3} = a_0 + a_1 E_t[\underbrace{y_{t+2}}_{\text{random at t}}] + a_2 E_t[\underbrace{y_{t+1}}_{\text{random at t}}] \quad (14a)$$

$$\text{Can be implemented using } \widehat{y}_{t+3} = \widehat{a}_0 = \widehat{a}_1 \widehat{y}_{t+2} + \widehat{a}_2 \widehat{y}_{t+1} \quad (14b)$$

Specific models: AR(2) - k period forecast

- k period ahead forecast

$$y_{t+k} = a_0 + a_1 y_{t+k-1} + a_2 y_{t+k-2} + \varepsilon_{t+k}$$

$$E_t y_{t+k} = a_0 + a_1 E_t[\underbrace{y_{t+k-1}}_{\text{random at t}}] + a_2 E_t[\underbrace{y_{t+k-2}}_{\text{random at t}}] \quad (15a)$$

Can be implemented using $\widehat{y}_{t+k} = \widehat{a}_0 = \widehat{a}_1 \widehat{y}_{t+k-1} + \widehat{a}_2 \widehat{y}_{t+k-2}$ (15b)

Specific models: MA(1) Model - 1 period forecast

3). Forecasting using MA(1) model

$$y_t = a_0 + \varepsilon_t + \beta_1 \varepsilon_{t-1}; \; E_{t-1}\varepsilon_t = 0 \text{ MA(1)}$$

- One period ahead forecast

$$y_{t+1} = a_0 + \varepsilon_{t+1} + \beta_1 \varepsilon_t$$

$$E_t y_{t+1} = a_0 + \underbrace{E_t[\underbrace{\varepsilon_{t+1}}_{\text{unknown at t}}]}_{=0 \text{ by assumption}} + \beta_1 E_t[\underbrace{\varepsilon_t}_{\text{known at t}}]$$

$$\longrightarrow E_t y_{t+1} = a_0 + \beta_1 \varepsilon_t \tag{16}$$

- Made feasible by $\widehat{y}_{t+1} = \widehat{a}_0 + \widehat{\beta}_1 \widehat{\varepsilon}_t$
- $\widehat{a}_0, \widehat{\beta}_1$ estimated by ML on eviews, matlab, R, stata etc. $\widehat{\varepsilon}_t$ is fitted (not forecasted) residual usually software gives you $\widehat{\varepsilon}_t$

Specific models: MA(1) Model - 1 period forecast cont.

- Or if $\widehat{\varepsilon}_1, ..., \widehat{\varepsilon}_T$ not given:
- Initialize by assuming:

$$\boxed{y_0 = \varepsilon_0 = \widehat{\varepsilon}_0 = 0} \tag{17}$$

- Then solve for $\widehat{\varepsilon}_1, ..., \widehat{\varepsilon}_T$ using the followings
  The fitted equations:

$$y_t = \widehat{a}_0 + \widehat{\varepsilon}_t + \widehat{\beta}_1 \widehat{\varepsilon}_{t-1} \tag{18}$$

$$(t = 1) \implies y_1 = \widehat{a}_0 + \widehat{\varepsilon}_1 + \widehat{\beta}_1 \underbrace{\widehat{\varepsilon}_0}_{=0 \text{ by initialization}} = \widehat{a}_0 + \widehat{\varepsilon}_1$$

$$\text{Now solve for } \widehat{\varepsilon}_1 : \boxed{\widehat{\varepsilon}_1 = \underbrace{y_1}_{\text{from data}} - \underbrace{\widehat{a}_0}_{\text{from software}}} \tag{19}$$

Specific models: MA(1) Model - 1 period forecast cont.

- Next

$$(t=2) \implies y_2 = \widehat{a}_0 + \widehat{\varepsilon}_2 + \widehat{\beta}_1 \widehat{\varepsilon}_1$$

Now solve for $\widehat{\varepsilon}_2$

$$\boxed{\widehat{\varepsilon}_2 = \underbrace{y_2}_{\text{from data}} - \underbrace{\widehat{a}_0 - \widehat{\beta}_1}_{\text{from software}} \times \underbrace{\widehat{\varepsilon}_1}_{\text{solved for in (19)}}} \tag{20}$$

- So, we assumed/set $\widehat{\varepsilon}_0 = 0$, then sovled for $\widehat{\varepsilon}_1$. And once we had $\widehat{\varepsilon}_1$ we could solve for $\widehat{\varepsilon}_2$. Now we have $\widehat{\varepsilon}_2$, we can solve for $\widehat{\varepsilon}_3$ and so on. Continue recursively using

$$(t=t) \implies y_t = \widehat{a}_0 + \widehat{\varepsilon}_t + \widehat{\beta}_1 \widehat{\varepsilon}_{t-1}$$

solve for

$$\boxed{\widehat{\varepsilon}_t = \underbrace{y_t}_{\text{from data}} - \underbrace{\widehat{a}_0 - \widehat{\beta}_1}_{\text{from software}} \times \underbrace{\widehat{\varepsilon}_{t-1}}_{\text{solved for this already}}} \tag{21}$$

Until we finally get to $\widehat{\varepsilon}_T = y_T - \widehat{a}_0 - \widehat{\beta}_1 \widehat{\varepsilon}_{T-1}$

Specific models: MA(1) Model - 2 period forecast

- Two period ahead forecast in MA(1)

$$y_t = a_0 + \varepsilon_t + \beta_1 \varepsilon_{t-1}, \;\; E_{t-1}\varepsilon_t = 0$$
$$\implies y_{t+2} = a_0 + \varepsilon_{t+2} + \beta_1 \varepsilon_{t+1}$$
$$\implies E_t y_{t+2} = a_0 + \underbrace{E_t[\varepsilon_{t+2}]}_{=0 \text{ by LIE}} + \beta_1 \underbrace{E_t[\varepsilon_{t+1}]}_{=0 \text{ by assumption}}$$

- So our forecast is:

$$\boxed{\widehat{y}_{t+2|t} = \widehat{a}_0}$$

Specific models: MA(1) Model - k period forecast and unconditional forecast

- K-period forecast for $k \geq 2$

$$
\begin{aligned}
y_{t+k} &= a_0 + \varepsilon_{t+k} + \beta_1 \varepsilon_{t+k-1} \\
E_t y_{t+k} &= a_0 + \underbrace{E_t \varepsilon_{t+k}}_{=0 \text{ for } k \geq 1} + \beta_1 \underbrace{E_t \varepsilon_{t+k-1}}_{=0 \text{ for } k \geq 2} = a_0
\end{aligned}
$$

- Unconditional forecast

$$
E[y_{t+k}] = a_0 + \underbrace{E[\varepsilon_{t+k}]}_{=0 \text{ by LIE}} + \beta_1 \underbrace{E[\varepsilon_{t+k-1}]}_{=0 \text{ by LIE}} = a_0
$$

Specific models: MA(1) Model - remarks and intuition

- In MA(1) model impact of shock lasts 2 periods
- So this period and next period share a common shock (ε_t)
- So this period's value helps forecast next period's value
- But this period does NOT share a shock with two periods later (ε_t does not show up in formula for y_{t+2})
- So y_t does not help forecast y_{t+2}
- And we revert back to the unconditional forecast

Specific models: MA(2) Model - 1 period forecast

4). MA(2) model:

$$y_t = a_0 + \varepsilon_t + \beta_1 \varepsilon_{t-1} + \beta_2 \varepsilon_{t-2}; \;\; E_{t-1}\varepsilon_t = 0$$

- One-period ahead forecast:

$$\begin{aligned} y_{t+1} &= a_0 + \varepsilon_{t+1} + \beta_1 \varepsilon_t + \beta_2 \varepsilon_{t-1} \\ \Longrightarrow E_t y_{t+1} &= a_0 + E_t[\varepsilon_{t+1}] + \beta_1 \varepsilon_t + \beta_2 \varepsilon_{t-1} \\ \Longrightarrow & \boxed{\widehat{y}_{t+1|t} = \widehat{a}_0 + \widehat{\beta}_1 \widehat{\varepsilon}_t + \widehat{\beta}_1 \widehat{\varepsilon}_{t-1}} \end{aligned}$$

Specific models: MA(2) Model - 2 and more period forecast

- Two-period ahead forecast:

$$y_{t+2} = a_0 + \underbrace{\varepsilon_{t+2}}_{\text{random at t}} + \beta_1 \underbrace{\varepsilon_{t+1}}_{\text{random at t}} + \beta_2 \underbrace{\varepsilon_t}_{\text{NR at t}}$$

$$\Longrightarrow E_t y_{t+2} = a_0 + \beta_2 \varepsilon_t \text{ by LIE and assumption}$$

$$\Longrightarrow \widehat{y}_{t+2} = \widehat{a}_0 + \widehat{\beta}_2 \widehat{\varepsilon}_t \text{ (2 period forecasts)}$$

- k-period ahead forecast ($k \geq 3$)):

$$y_{t+k} = a_0 + \underbrace{\varepsilon_{t+k}}_{\text{random at t}} + \beta_1 \underbrace{\varepsilon_{t+k-1}}_{\text{random at t}} + \beta_2 \underbrace{\varepsilon_{t+k-2}}_{\text{random at t}}$$

$$\Longrightarrow E_t y_{t+k} = a_0 \text{ (3-periods forecasts)}$$

Specific models: ARMA(1,1) Model - 1 period forecast

5). ARMA(1,1) model:

$$y_t = a_0 + a_1 y_{t-1} + \varepsilon_t + \beta_1 \varepsilon_{t-1};\ \ E_{t-1}\varepsilon_t = 0$$

- One-period ahead forecast:

$$y_{t+1} = a_0 + a_1 y_t + \varepsilon_{t+1} + \beta_1 \varepsilon_t$$

$$\Longrightarrow E_t y_{t+1} = a_0 + a_1 y_t + \underbrace{E_t \varepsilon_{t+1}}_{=0} + \beta_1 \underbrace{\varepsilon_t}_{\text{NR at t}}$$

$$= a_0 + a_1 y_t + \beta_1 \varepsilon_t$$

$$\boxed{\widehat{y}_{t+1|t} = \widehat{a}_0 + \widehat{a}_1 y_t + \widehat{\beta}_1 \widehat{\varepsilon}_t}$$

- Two-period ahead forecast:

$$y_{t+2} = a_0 + a_1 y_{t+1} + \varepsilon_{t+2} + \beta_1 \varepsilon_{t+1}$$

$$E_t y_{t+2} = a_0 + a_1 E_t[y_{t+1}] + \underbrace{E_t[\varepsilon_{t+2}]}_{=0} + \beta_1 \underbrace{E_t[\varepsilon_{t+1}]}_{=0}$$

$$\Longrightarrow \boxed{\widehat{y}_{t+2|t} = \widehat{a}_0 + \widehat{a}_1 \widehat{y}_{t+1|t}}$$

Specific models: ARMA(1,1) Model - k period forecast

- k-period ahead forecast ($k > 2$):

$$
\begin{aligned}
&y_{t+k} = a_0 + a_1 y_{t+k-1} + \varepsilon_{t+k} + \beta_1 \varepsilon_{t+k-1} \\
&\implies E_t y_{t+k} = a_0 + a_1 E_t[y_{t+k-1}] + \underbrace{E_t[\varepsilon_{t+k}]}_{=0} + \beta_1 \underbrace{E_t[\varepsilon_{t+k-1}]}_{=0} \\
&= a_0 + a_1 E_t[y_{t+k-1}] \\
&\boxed{\widehat{y}_{t+k|t} = \widehat{a}_0 + \widehat{a}_1 \widehat{y}_{t+k-1|t}} \text{ for } k > 2
\end{aligned}
$$

- So, in summary

$$
\widehat{y}_{t+k|t} = \begin{cases} \widehat{a}_0 + \widehat{a}_1 y_t + \widehat{\beta}_1 \widehat{\varepsilon}_t \ , & \text{for k=1} \\ \widehat{a}_0 + \widehat{a}_1 \widehat{y}_{t+k-1|t} \ , & \text{for k} \geq 2 \end{cases}
$$

6). ARMA(p,q) model:

$$y_t = a_0 + a_1 y_{t-1} + ... + a_p y_{t-p} + \varepsilon_t + \beta_1 \varepsilon_{t-1} + ... + \beta_q \varepsilon_{t-q}$$

- k-period ahead forecast

$$y_{t+k} = a_0 + a_1 y_{t+k-1} + ... + a_p y_{t+k-p} + \varepsilon_{t+K} + \beta_1 \varepsilon_{t+k-1} + ...$$
$$+ \beta_q \varepsilon_{t+k-q}$$
$$\Longrightarrow E_t y_{t+k} = a_0 + a_1 E_t[y_{t+k-1}] + ... + a_p E_t[y_{t+k-p}] + \underbrace{E_t \varepsilon_{t+k}}_{=0}$$
$$+ \beta_1 E_t \varepsilon_{t+k-1} + ... + \beta_q E_t \varepsilon_{t+k-q}$$
$$\Longrightarrow E_t y_{t+k} = a_0 + \sum_{i=1}^{p} a_i E_t y_{t+k-i} + \sum_{j=1}^{q} \beta_j E_t \varepsilon_{t+k-j}$$

Specific models: ARMA(p,q) Model - k period forecast cont.

- Note that

$$E_t y_{t+k-i} = \begin{cases} y_{t+k-i}\ , \text{ for } t+k-i \leq t \iff i \geq k \\ E_t y_{t+k-i}\ , \text{ for } i < k \end{cases}$$

$$E_t \varepsilon_{t+k-j} = \begin{cases} \varepsilon_{t+k-j}\ , \text{ for } t+k-j \leq t \iff j \geq k \\ 0\ , \text{ for } j < k \end{cases}$$

- Therefore,

$$E_t y_{t+k} = \begin{cases} a_0 + \sum_{i=1}^{k-1} E_t y_{t+k-i} + \sum_{i=k}^{p} a_j y_{t+k-i} \\ + \sum_{j=k}^{q} \beta_j \varepsilon_{t+k-j},\ \underline{\text{for}}\ k \leq p\ \&k \leq q \\ a_0 + \sum_{i=1}^{p} a_i E_t y_{t+k-i},\ \underline{\text{for}}\ k > p\ \&k > q \\ a_0 + \sum_{i=1}^{k-1} a_i E_t y_{t+k-i} + \sum_{j=k}^{p} a_j y_{t+k-j},\ \underline{\text{for}}\ p \geq k \geq q \\ a_0 + \sum_{i=1}^{p} a_i E_t y_{t+k-i} + \sum_{j=k}^{q} \beta_j \varepsilon_{t+k-j},\ \underline{\text{for}}\ q \geq k \geq p \end{cases}$$

Specific models: ARMA(p,q) Model - k period forecast cont.

- So in practice using

$$\widehat{y}_{t+k|t} = \begin{cases} \widehat{a}_0 + \sum_{i=1}^{k-1} \widehat{a}_i \widehat{y}_{t+k-i|t} + \sum_{i=k}^{p} \widehat{a}_i y_{t+k-i} \\ + \sum_{j=k}^{q} \widehat{\beta}_j \widehat{\varepsilon}_{t+k-j}, \ \underline{\text{for}}\ k \leq p \ \& k \leq q \\ \widehat{a}_0 + \sum_{i=1}^{p} \widehat{a}_i \widehat{y}_{t+k-i|t}, \ \underline{\text{for}}\ k > p \ \& k > q \\ \widehat{a}_0 + \sum_{i=1}^{k-1} \widehat{a}_i \widehat{y}_{t+k-i|t} + \sum_{i=k}^{p} \widehat{a}_i y_{t+k-i}, \ \underline{\text{for}}\ p \geq k \geq q \\ \widehat{a}_0 + \sum_{i=1}^{p} \widehat{a}_i \widehat{y}_{t+k-i|t} + \sum_{j=k}^{q} \widehat{\beta}_j \varepsilon_{t+k-j}, \ \underline{\text{for}}\ q \geq k \geq p \end{cases}$$

Specific models: ARMA(p,q) Model - remark

- Rather than trying to memorize or reproduce this general case, it is probably better practice to derive the forecast for a number of specific values of q & q.
- As practice, I suggest working out the forecasts yourself the following models: ARMA(2,1), ARMA(1,2) & ARMA(2,2)
- If you want additional practice, try ARMA(2,3) & ARMA(3,2)

Constructing out of sample forecasts

- Approach 1: Estimation and forecast samples
- Approach 2: recursive forecasting
- Approach 3: rolling samples

In sample prediction versus out-of-sample forecasts

- Suppose you estimate the regression:

$$y_t = \beta_0 + \beta x_{t-1} + \varepsilon_t$$

  using data $x_1, x_2, ..., x_{T-1}$ & $y_2, y_3, ..., y_T$.
- Let $\widehat{y}_t = \widehat{\beta}_0 + \widehat{\beta}_1 x_{t-1}$
- $\widehat{y}_t = \begin{cases} \text{in sample prediction for } t \leq T \\ \text{out-of-sample forecast for } t > T \end{cases}$
- Key point: cannot data that you are trying to forecast to fit the mode

Generating a sample out of sample forecasts

- May want to evaluate forecast method
  - ∗ How well are we forecasting?
  - ∗ IS another method better?
- To answer this, essentially want to measure, estimate, or test forecast accuracy
- To do this, we need a sample of forecasts and forecast errors large enough for meaningful statistical inference

Approach 1: Estimation and forecast samples

- <u>Approach 1: Estimation and Forecast Samples</u>
  - Full sample: $S = \{y_1, y_2, ..., y_{T_1}\}$
  - Rather than estimate model using full sample S, split it into 2 subsamples
  - Estimation sample $S_1 = \{y_1, y_2, ..., y_{T_1}\}$
    Forecast sample $S_2 = \{y_{T_1+1}, y_{T_1+2}, ..., y_T\}$
  - Estimate the model using only the data in the estimation sample
  - Use the resulting estimates to produce forecasts in the forecast sample

Approach 1: example

- Example: $(y_t - \bar{y}) = a_1(y_{t-1} - \bar{y}) + \varepsilon_t$
  I have data on $y_1, y_2, ..., y_{200}$

$$S = \{y_1, y_2, ..., y_{200}\}$$
$$S_1\{y_1, y_2, ..., y_{100}\}$$
$$S_2 = \{y_{101}, y_{102}, ..., y_{200}\}$$

- Estimation (use S_1):

$$\left.\begin{aligned} \bar{y} &= \frac{1}{100}\sum_{t=1}^{100} y_t \\ \widehat{a}_1 &= \frac{\sum_{t=2}^{100}(y_{t-1} - \bar{y})(y_t - \bar{y})}{\sum_{t=2}^{100}(y_{t-1} - \bar{y})^2} \end{aligned}\right\} \tag{22}$$

- Forecast $y_{101}, y_{102}, ..., y_{200}$

Approach 1: example cont.

- Note that

$$
\begin{aligned}
(y_t - \bar{y}) &= \widehat{a}_1(y_{t-1} - \bar{y}) + \widehat{\varepsilon}_t \\
y_t &= \bar{y} + \widehat{a}_1(y_{t-1} - \bar{y}) + \widehat{\varepsilon}_t \\
y_t &= \bar{y}(1 - \widehat{a}_1) + \widehat{a}_1 y_{t-1} + \widehat{\varepsilon}_t \\
\implies \widehat{y}_{t+1|t} &= (1 - \widehat{a}_1)\bar{y} + \widehat{a}_1 y_{t-1}
\end{aligned}
$$

- Hence,

$$
\begin{aligned}
&\widehat{y}_{101|100} = (1 - \widehat{a}_1\bar{y}) + \widehat{a}_1 y_{100}; \;\; e_{101|100} = \widehat{y}_{101|100} - y_{101} \\
&\widehat{y}_{102|101} = (1 - \widehat{a}_1\bar{y}) + \widehat{a}_1 y_{101}; \;\; e_{102|101} = \widehat{y}_{102|101} - y_{102} \\
&\ldots \\
&\widehat{y}_{200|199} = (1 - \widehat{a}_1\bar{y}) + \widehat{a}_1 y_{199}; \;\; e_{200|199} = \widehat{y}_{200|199} - y_{200}
\end{aligned}
$$

- This gives us a sample of:
  - $*$ 100 forecasts: $\{\widehat{y}_{101|100}, \widehat{y}_{101|101}, \ldots, \widehat{y}_{200|199}\}$
  - $*$ And 100 forecast errors: $\{e_{101|100}, e_{102|101}, \ldots, e_{200|199}\}$

Approach 1: example pseudo code

- What does the program to this look like?
  - Load data=datafile;
  - **Estimation**:
    S_1=data[row 1 to 100];
    =mean(S$_1$);
    $\widehat{a}_1$=AR1(S$_1$);
  - **Forecasts**:
    S_2 =data[row 101 to 200];
    Store Forecasts = 1 by 100 matrix of zeros;
    Store Errors = 1 by 100 matrix of zeros;
    t=100;
    do until t>199;
      $\widehat{y}_{t+1|t} = (1 - \widehat{a}_1)\bar{y} + \widehat{a}_1 y_t$;
      $e_{t+1|t} = \widehat{y}_{t+1|t}$;
      Store Forecast[row t-99]=$\widehat{y}_{t+1|t}$;
      Store Error[row t-99]=$e_{t+1|t}$;
      t=t+1;
    end loop
  - Then our forecasts would be stored in the matrices: Store Forecast & Store Error & we could use to evaluate our forecasts.

Approach 2: recursive forecasting

- Approach 2: Recursive Forecasting
  - $S = \{y_1, y_2, ..., y_T\}$ full sample
  - Now break into:
    - $S_1 = \{y_1, y_2, ..., y_{T_1}\}$= training sample
    - $S_2 = \{y_{T_1+1}, y_{T_1+2}, ..., y_T\}$= recursive forecast sample
  - We forecast $y_{T_1+1}, y_{T_1+2}, ..., y_T$
  - To forecast y_{T_1+1}, we
    1. Estimate model using $y_1, y_2, ..., y_{T_1}$
    2. Use estimated model to forecast y_{T_1+1}
  - So far, that's no different than approach 1
  - However, we now re-estimate our model using data up to y_{T_1+1} in order to forecast y_{T_1+2}
  - And we estimate using data up to y_{T_1+2} to forecast y_{T_1+3}...
  - Etc

Approach 2: example

- Example: $(y_t - \bar{y}) = a_1(y_{t-1} - \bar{y}) + \varepsilon_t$
  sample $y_1, y_2, ..., y_{200}$
  $S_1 = \{y_1, y_2, ..., y_{100}\}$ $S_2 = \{y_{101}, y_{102}, ..., y_{200}\}$
- Step 1: estimate using $y_1, y_2, ..., y_{100}$
  $\bar{y}_{100} = \frac{1}{100}\sum_{t=1}^{100} y_t$
  $\widehat{a}_{1,100} = \frac{\sum_{t=2}^{100}(y_{t-1}-\bar{y}_{100})(y_t-\bar{y}_{100})}{\sum_{t=2}^{100}(y_{t-1}-\bar{y}_{100})^2}$
- Step 2: forecast y_{101}
  $\widehat{y}_{101|100} = (1 - \widehat{a}_{1,100})\bar{y}_{100} + \widehat{a}_{1,100}y_{100}$
- Step 3: re-estimate using $y_1, y_2, ..., y_{101}$
  $\bar{y}_{101} = \frac{1}{101}\sum_{t=1}^{101} y_t$
  $\widehat{a}_{1,101} = \frac{\sum_{t=2}^{101}(y_{t-1}-\bar{y}_{101})(y_t-\bar{y}_{101})}{\sum_{t=2}^{101}(y_{t-1}-\bar{y}_{101})^2}$
- Step 4: forecast y_{102} $\widehat{y}_{102|101} = (1 - \widehat{a}_{1,101})\bar{y}_{101} + \widehat{a}_{1,101}y_{101}$
- ...

Approach 2: example cont.

- Generally, re-estimate using $y_1, y_2, ..., y_t$ $(100 \leq t \leq 199)$
  $\bar{y}_t = \frac{1}{s} \sum_{s=1}^{t} y_s$
  $\widehat{a}_{1,t} = \frac{\sum_{s=2}^{t} (y_{s-1} - \bar{y}_t)(y_s - \bar{y}_t)}{\sum_{s=2}^{t} (y_{s-1} - \bar{y}_t)^2}$
  And forecast y_{t+1}:
  $\widehat{y}_{t+1|t} = (1 - \widehat{a}_{1,t})\bar{y}_t + \widehat{a}_{1,t} y_t$
- Again, this produces a sample of forecasts:
  $\{\widehat{y}_{101|100}, \widehat{y}_{102|101}, ..., \widehat{y}_{200|199}\}$ along with a corresponding sample of forecast errors

Approach 2: example pseudo code

- What might the program code for this look like?
  - `load data =datafile;`
  - **Estimation and Forecast**:

```
Store Error = 1 by 100 matrix of zeros;
Store Forecast = 1 by 100 matrix of zeros;
t=100;
do until t>199;
 S=data[rpw 1 to t];
```

$\bar{y}_t$`=mean(S);`

$\widehat{a}_{1,t}$ `=AR1(S);`

$\widehat{y}_{t+1|t} = (1-\widehat{a}_{1,t})\bar{y} + \widehat{a}_{1,t}\underbrace{\texttt{S[row t]}}_{y_t};$

`Store Forecast[row t-99]=`$\widehat{y}_{t+1|t}$;

$\texttt{error}_{t+1} = \widehat{y}_{t+1|t}$-$\underbrace{\texttt{data[row t+1]}}_{y_{t+1}}$

`Store Error[row t-99]=`$\texttt{error}_{t+1}$

`t=t+1; end loop;`

  - Again, we have a sample of forecats and forecast errors stored in the matrices "`Store Forecast`" & "`Store Error`"

Approach 3: rolling samples

- If we think there are occasional structural breaks, we may not want to use entire history for estimation since that might include data generated by an old, pre-break, model
- To construct a rolling sample, we choose a fixed sub-sample size (say η)
- Then, we always use the most recent η periods to re-estimate the model
- I.e. to forecast y_{t+1}, we estimate the forecast model using sub-sample of data: $\{y_{t-\eta+1}, y_{t-\eta+2}, ..., y_t\}$

Approach 3: example

- Example: $(y_t - \bar{y}) = a_1(y_{t-1} - \bar{y}) + \varepsilon_t,\ y_1, y_2, .., y_{200}$
- Say we choose $\eta = 50$
- Step 1: estimate using $y_1, y_2, ..., y_{50}$
  $\bar{y}_{1,1:50} = \frac{1}{50}\sum_{s=1}^{50} y_s$
  $\widehat{a}_{1,1:50} = \frac{\sum_{s=2}^{50}(y_{s-1}-\bar{y}_{1:50})(y_s-\bar{y}_{1:50})}{\sum_{s=2}^{50}(y_{s-1}-\bar{y}_{1:50})^2}$
- Step 2: forecast y_{51}
  $\widehat{y}_{51|50} = (1 - \widehat{a}_{1,1:50})\bar{y}_{1:50} + \widehat{a}_{1,1:50}y_{50}$
- Step 3: re-estimate using $y_2, y_3, ..., y_{51}$
  $\bar{y}_{2:51} = \frac{1}{50}\sum_{s=2}^{51} y_s$
  $\widehat{a}_{1,2:51} = \frac{\sum_{s=3}^{51}(y_{s-1}-\bar{y}_{2:51})(y_s-\bar{y}_{2:51})}{\sum_{s=3}^{51}(y_{s-1}-\bar{y}_{2:51})^2}$
- ...
  etc

LIE detail

LIE

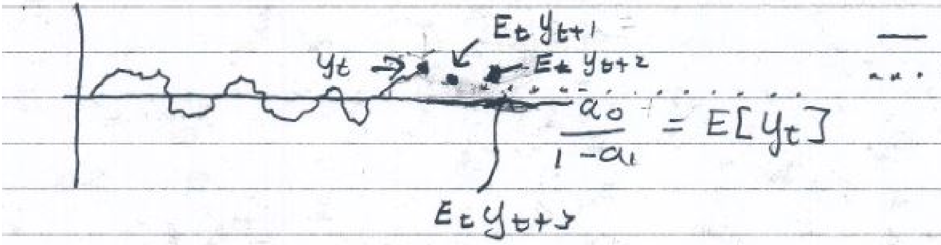
Law of Iterative Expectation (L.I.E)

$$\underbrace{E_t(y)}_{\text{today's forecast}} = \underbrace{E_t\{\underbrace{E_{t+k}[y]}_{\text{tomorrow's forecast}}\}}_{\substack{\text{today's forecast} \\ \text{for tomorrow's forecast}}} \text{, any } k > 0$$

Since, intuitively $E = E_{-\infty}$ another form of the L.I.E is :

$$E[y] = E[E_t[y]]$$

back



You should include time plots like this when submitting your weekly forecasts.

Back